OU_156340

THE COORDINATE METHOD

THE POCKET MATHEMATICAL LIBRARY
JACOB T. SCHWARTZ, *Editor*

PRIMERS:

1. THE COORDINATE METHOD
by I. M. Gelfand et al.

2. FUNCTIONS AND GRAPHS
by I. M. Gelfand et al.

WORKBOOKS:

1. SEQUENCES AND COMBINATORIAL PROBLEMS
by S. I. Gelfand et al.

2. LEARN LIMITS THROUGH PROBLEMS!
by S. I. Gelfand et al.

3. MATHEMATICAL PROBLEMS: AN ANTHOLOGY
by E. B. Dynkin et al.

COURSES:

1. LIMITS AND CONTINUITY
by P. P. Korovkin

2. DIFFERENTIATION
by P. P. Korovkin

THE COORDINATE METHOD

BY

I. M. GELFAND, E. G. GLAGOLEVA
AND A. A. KIRILLOV

Revised English edition
Translated and Edited by
RICHARD A. SILVERMAN

Contents

Preface

This book, the first primer in The Pocket Mathematical Library, is devoted to the coordinate method. It requires nothing more than a knowledge of elementary high school mathematics. However, the book is written for systematic study rather than for quick reading. For this reason, you may find the going a bit hard. To make things easier, certain road signs have been put in the margins:

The parking sign indicates places containing information indispensable for further reading, e.g., definitions, formulas, etc. You should linger at such places, mulling things over and making sure they stick in your mind.

The grade sign indicates more difficult material which can be omitted on a first reading if it appears between asterisks.

Pay special attention to the danger sign. It often appears near a passage which at first glance seems easy and straightforward. However, failure to think these passages through can lead to serious errors later.

The importance of solving the problems cannot be exaggerated. Good luck!

Introduction

In an account of the launching of a new satellite, you might well read that "the actual orbit of the satellite is close to the calculated orbit". Ask yourself how an orbit, which is a curve in space, can be *calculated*, i.e., expressed in numbers. This can only be done if geometric notions can be translated into numerical language, a possibility which depends in the first instance on being able to specify the position of a point in space (or in a plane or on the earth's surface) by means of numbers.

By the *coordinate method* we mean the technique of using numbers (or other symbols) to specify the position of a point (or of an object), and the numbers themselves are called the *coordinates* of the point.

For example, you already know how geographical coordinates are used to specify the position of a point on the earth's surface. In fact, every point on the earth's surface has two coordinates, its latitude and its longitude. To specify the position of a point in space, we need three numbers rather than two. Thus to specify the position of a satellite, we can give its height above the earth's surface as well as the latitude and longitude of the point of the earth's surface directly under the satellite. However, suppose we know the orbit of the satellite, i.e., the curve along which it moves. Then to specify the position of the satellite we need only one number, for example the distance traversed by the satellite measured from some point of its orbit. Thus curves are "one-dimensional" and surfaces are "two-dimensional," while space is "three-dimensional." In other words, we need one coordinate to specify the position

of a point along a curve, two coordinates to specify position along a surface and three to specify spatial position.

In just the same way, the coordinate method can be used to specify the position of a train along a railroad track by giving the number of the nearest milestone, or of an automobile along a road by giving the distance to a key destination (e.g., New York along the New Jersey Turnpike).
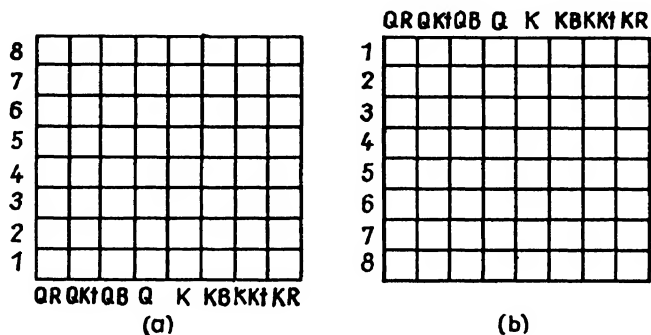
Fig. 1

Coordinates of a special kind are used in chess, where the positions of the pieces on the board are indicated by letters and numbers. Here the vertical rows are labelled by symbols indicating the piece which occupies the first square in the row at the beginning of the game, while the horizontal rows are labelled by the numbers 1 through 8 with each player counting from his side of the board. Thus Figure 1a shows a chess board as seen from White's side, while Figure 1b shows the same board as seen from Black's side. In both cases, we read "Q, QB, QKt, QR" as "Queen, Queen's Bishop, Queen's Knight, Queen's Rook" and "K, KB, KKt, KR" as "King, King's Bishop, King's Knight, King's Rook." Use of these coordinates makes it possible to record games or to play chess by mail without bothering to make drawings showing the positions of the pieces on the board. For example, if a hyphen means

"moves to," then the shortest chess game, called the "fool's mate," in which Black checkmates White in two moves, takes the following form:

White	*Black*
1. P–KB4	1. P–K3
2. P–KKt4	2. Q–KR5

In mathematics, coordinates can be used to describe the position of any point, whether it be a point in space, in the plane or on a curve. This makes it possible to "quantify" all kinds of geometric concepts, describing them in terms of numbers. An example of such a description is given in Problem 1, p. 11.

The coordinate method is particularly important in that it allows modern electronic calculating machines to be used to solve a great variety of geometric problems and to investigate geometric objects and relationships.

PART 1

CLASSICAL
COORDINATE GEOMETRY

CHAPTER 1

Coordinates on a Line

1. The Coordinate Line

We begin our study of coordinates by analyzing the simple problem of how to specify the position of a point on a line. This is done as follows: On the line we choose an *origin* (some point O), a *unit of length* (a line segment e) and a *direction* which can be regarded as *positive* (indicated by an arrow in Figure 2).

Fig. 2

A straight line equipped with an origin, a unit of length and a positive direction is called a *coordinate line* (or *axis*).

To specify the position of a point on a coordinate line, it is only necessary to give a single number, for example $+5$. This means that the point so labelled lies at a distance of 5 units from the origin in the positive direction.

The number specifying the position of a point on a coordinate line is called the *coordinate* of the point on the axis.

The coordinate of a point on a coordinate line equals the distance of the point from the origin, expressed in the chosen units of measurement and taken with a plus sign if the point lies in the positive direction from the origin and with the minus sign if the point lies in the negative direction. The origin is often called the *origin of coordinates*. The coordinate of the origin (the point O) equals zero.

Notation like $M(-7)$, $A(x)$, etc. is often used. The first means the point M with coordinate -7, the second the point A with coordinate x. For brevity, one often says "the point -7," "the point x," and so on.

3
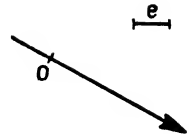
By introducing coordinates, we establish a correspondence between numbers and points of a straight line. In doing so, we exploit the following remarkable fact: There is a unique number corresponding to each point of the line and a unique point of the line corresponding to each number.

Next we define a key term: A correspondence between two sets is said to be *one-to-one* if a unique element of the second set corresponds to each element of the first set and if (under the same correspondence) a unique element of the first set corresponds to each element of the second set. Thus the correspondences shown in Examples a and c of Figure 3 are one-to-one, but not the correspondences shown in Examples b and d. At first glance, the problem of establishing a one-to-one correspondence between numbers and points of the line seems to be quite a simple matter.
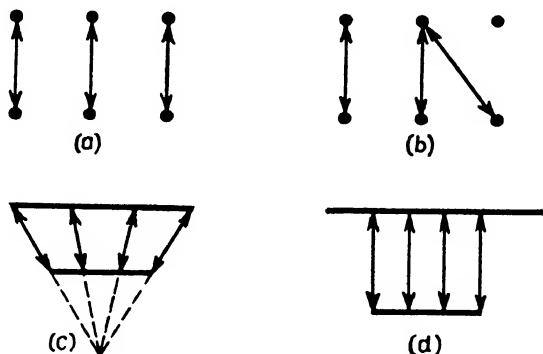
Fig. 3

However, when mathematicians began to think things over carefully, it turned out that an extensive and subtle theory was needed to clarify the concepts involved. For example, the answers to the two seemingly simple questions "What is a number?" and "What is a point?" both turn out to be difficult. Indeed these are questions pertaining to the foundations of geometry and the axiomatization of arithmetic.

Regardless of the fact that the problem of specifying the position of a point on a line is very simple, you must analyze it very carefully in order to become accustomed to think of numerical relationships in geometrical terms and conversely.

If you have properly understood Section 1, then you will have no difficulty with the following set of problems. On the other hand, if you find the problems hard, then you must have failed to understand some key point. If so, go back and read the section again!

PROBLEMS

1. a) Plot the points $A(-2)$, $B(\frac{13}{3})$ and $C(0)$ on a coordinate line.

b) Plot the point $M(2)$ on a coordinate line. Find the two points A and B at a distance of three units from M. What are the coordinates of the points A and B?

2. a) Suppose the point $A(a)$ lies to the right of the point $B(b)$.[1] Which of the numbers is larger, a or b?

b) Without plotting the following pairs of points on a coordinate line, ascertain which lies to the right of the other:

a) $A(-3)$, $B(-4)$; b) $A(3)$, $B(4)$; c) $A(-3)$, $B(4)$;

d) $A(3)$, $B(-4)$.

3. Which of the points $A(a)$, $B(-a)$ lies to the right of the other?

Answer. Indeterminate. A lies to the right of B if a is positive and to to the left of B if a is negative.

4. Which of the following pairs of points lies to the right of the other:

a) $M(x)$, $N(2x)$; b) $A(c)$, $B(c + 2)$; c) $A(x)$, $B(x - a)$;

d) $A(x)$, $B(x^2)$.

Answer. c) A lies to the right of B if a is greater than zero and to the left of B if a is less than zero. A and B coincide if $a = 0$.

5. Plot the points $A(-5)$ and $B(7)$ on a coordinate line. Find the coordinate of the midpoint of the segment AB.

6. Using a red pencil, plot the points whose coordinates are

a) Integers; b) Positive numbers.

7. Find all points x on the coordinate line such that

a) $x < 2$; b) $x \geqslant 5$; c) $2 < x < 5$; d) $-\frac{13}{4} \leqslant x \leqslant 0$.

---

1. Here and henceforth, we assume that the axis is horizontal and that the positive direction is the direction from left to right.

2. The Absolute Value of a Number

By the *absolute value* (or *modulus*) of the number x is meant the distance of the point $A(x)$ from the origin.

The absolute value of x is indicated by vertical lines. Thus $|x|$ is the absolute value of x. For example,

$$|-3| = 3, \quad |\tfrac{1}{2}| = \tfrac{1}{2}.$$

It follows that

$$|x| = x \text{ if } x > 0,$$

$$|x| = -x \text{ if } x < 0,$$

$$|x| = 0 \text{ if } x = 0.$$

Since the points a and $-a$ lie at the same distance from the origin, the numbers a and $-a$ have the same absolute value:

$$|a| = |-a|.$$

PROBLEMS

1. What are the possible values of the expression

$$\frac{|x|}{x} ?$$

2. Write the following expressions without the absolute value signs:

  a) $|a^2|$; b) $|a - b|$ if $a > b$; c) $|a - b|$ if $a < b$; d) $|-a|$ if a is negative.

3. Given that $|x - 3| = x - 3$, find x.

4. Find the points on the coordinate line such that

  a) $|x| = 2$; b) $|x| > 3$; c) $|x| \leqslant 5$; d) $3 < |x| < 5$;

  e) $|x - 2| = 2 - x$.

Hint. b) If x is positive, then $|x| = x$ and hence $x > 3$; if x is negative, then $|x| = -x$, so that the inequality $-x > 3$ implies $x < -3$.

Answer. b) All points to the left of the point -3 and all points to the right of the point 3. The answer can be obtained more quickly by noting that $|x|$ is the distance of the point x from the origin. It is then clear that the points such that $|x| > 3$ are those whose distance from the origin exceeds 3, as can be seen by drawing a figure.

5. Solve the equation

   a) $|x - 2| = 3$;   b) $|x + 1| + |x + 2| = 1$.

Answer. b) The equation has infinitely many solutions which together fill up the interval $-2 \leqslant x \leqslant -1$. In other words, the equation is satisfied by any number greater than or exceeding -2 or by any number less than or equal to -1.

3. The Distance Between Two Points

We now derive a general formula for the distance between two points on the coordinate line:

Problem. Find the distance $\varrho\,(A, B)$ *between two given points* $A(x_1)$ *and* $B(x_2)$.[2]

Solution. We must analyze all possible relative positions of the origin O and the points A and B. There are six possible cases. First we consider the three cases where B lies to the right of A, as shown in Figures 4a, 4b and 4c. In the first case (Fig. 4a), the distance $\varrho\,(A, B)$ equals the difference between the distances of the points B and A from the origin. Since x_1 and x_2 are positive in this case, we have

$$\varrho\,(A, B) = x_2 - x_1.$$

In the second case (Fig. 4b), the distance equals the sum of the distances of the points B and A from the origin, i.e.,
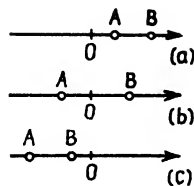
$$\varrho\,(A, B) = x_2 - x_1$$

Fig. 4

---

2. The Greek letter ϱ ("rho") is ordinarily used to denote distance. By the same token, $\varrho\,(A, B)$ is used to denote the distance between the points A and B.
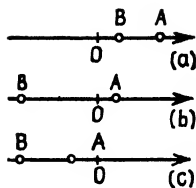
Fig. 5

as before, since in this case x_2 is positive and x_1 is negative. In the third case (Fig. 4c), the distance is still given by the same formula (why?).

The three other cases shown in Figure 5 differ from the first three in that the roles of the points A and B are reversed. In each of these cases, it can be verified that the distance between the points A and B equals

$$\varrho(A, B) = x_1 - x_2.$$

Thus in all cases where $x_2 > x_1$, the distance $\varrho(A, B)$ equals $x_2 - x_1$, while in all cases where $x_1 > x_2$, the distance equals $x_1 - x_2$. Recalling the definition of the absolute value, we can write a single formula comprising all six cases:

$$\varrho(A, B) = |x_2 - x_1|.$$

This formula can also be written in the form

$$\varrho(A, B) = |x_1 - x_2|.$$

At the risk of seeming pedantic, we must also consider the case where $x_2 = x_1$, i.e., where the points A and B coincide. It is clear that in this case

$$\varrho(A, B) = |x_2 - x_1|$$

also holds! Our problem is now completely solved.

PROBLEMS

1. Find the distance between the points

a) $A(-7)$ and $B(-2)$;   b) $A(-\frac{7}{2})$ and $B(-9)$.

2. Find the points x on the coordinate line such that

a) $\varrho(x, 7) < 3$;   b) $|x - 2| > 1$;   c) $|x + 3| = 3$.

3. Given two points $A(x_1)$ and $B(x_2)$ on the coordinate line, find the coordinate of the midpoint of the segment AB.

Remark. Consider all possible positions of the points A and B on the coordinate line or else write a solution valid for all cases at once.

4. Find the coordinate of the point on the coordinate line which is three times closer to the point $A(-9)$ than to the point $B(-3)$.

5. Solve Problem 5, p. 7 using the concept of the distance between two points.

6. Solve the following equations:

a) $|x + 3| + |x - 1| = 5$;

b) $|x + 3| + |x - 1| = 4$;

c) $|x + 3| + |x - 1| = 3$;

d) $|x + 3| - |x - 1| = 5$;

e) $|x + 3| - |x - 1| = 4$;

f) $|x + 3| - |x - 1| = 3$.

CHAPTER 2

Coordinates in a Plane

4. The Cartesian Plane

Next we consider the problem of specifying the coordinates of a point in the plane. First we draw two perpendicular coordinate axes (lines) in the plane, one called the *axis of abscissas* or the *x-axis*, denoted by Ox, and the other called the *axis of ordinates* or the *y-axis*, denoted by Oy.

The direction of the axes is usually chosen in such a way that a rotation through 90° in the counterclockwise direction causes the positive *x*-axis to coincide with the positive *y*-axis (see Fig. 6 a). The point of intersection of the axes is called the *origin of coordinates*, denoted by the letter O, and serves as the origin for both of the coordinate axes Ox and Oy. As a rule, the units of length are chosen to be the same on both lines.

Given any point M in the plane, we now drop perpendiculars from M to the x and y-axes, intersecting the axes in points M_1 and M_2 (see Fig. 6 b), called the *projections* of M onto the x and y-axes. M_1 is a point of the axis Ox and as such has a definite coordinate x. Similarly, M_2 is a point of the axis Oy and as such has a definite coordinate y.

Thus to every point M in the plane there corresponds a pair of numbers x and y, called the *rectangular Cartesian coordinates* of M. The number x is called the *abscissa* of M, and the number y is called its *ordinate*. The plane itself, equipped with these coordinates, is often called the *Cartesian plane*.
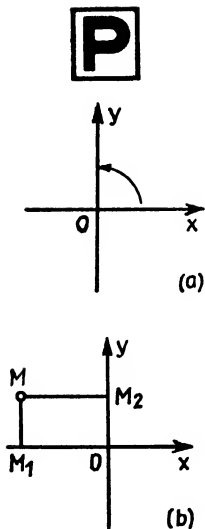
Fig. 6

Conversely, given any pair of numbers x and y, there is a point of the plane with x as abscissa and y as ordinate.

This establishes a one-to-one correspondence[1] between the points of the plane and pairs of numbers x and y, written in in a definite order (first x, then y).

Thus by the *rectangular Cartesian coordinates* x, y of a point M in the plane, we mean the coordinates of the projections of M onto the axes Ox and Oy, regarded as points of these axes.

The point M with coordinates x, y is usually denoted by $M(x, y)$ or simply (x, y), with the abscissa x written first and then the ordinate y. For brevity, we sometimes say "the point $(3, -8)$," for example, rather than "the point with coordinates $3, -8$."

The coordinate axes divide the plane into four regions called *quadrants*. The first quadrant lies between the positive x-axis and the positive y-axis, and the other quadrants are labelled in order in the counterclockwise direction (see Fig. 7).

To master coordinates in the plane, you should now solve the following problems:

Fig. 7

PROBLEMS

We begin with a few easy exercises.

1. What word is spelled by the points

$(-9,0)$, $(-5,1)$, $(2,4)$, $(8,3)$, $(5,0)$, $(6,2)$, $(2,3)$, $(-3,2)$, $(-4,2)$, $(4,4)$, $(5,3)$, $(7,2)$, $(-4,4)$, $(2,0)$, $(5,4)$, $(0,4)$, $(-6,3)$, $(8,1)$, $(3,4)$, $(-4,3)$, $(-1,3)$, $(-1,2)$, $(8,2)$, $(5,1)$, $(2,2)$, $(-1,4)$, $(-9,4)$, $(-5,3)$, $(5,2)$, $(1,4)$, $(-1,0)$, $(-2,4)$, $(-9,3)$, $(-5,4)$, $(-1,1)$, $(-4,0)$, $(-7,2)$, $(-8,3)$, $(-3,4)$, $(-2,2)$, $(-5,2)$, $(-5,0)$, $(-4,1)$, $(-9,1)$, $(-9,2)$, $(2,1)$, $(8,0)$, $(8,4)$?
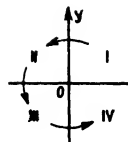
---

1. A one-to-one correspondence between the points of the plane and pairs of numbers is a correspondence which associates a unique pair of numbers with every point and a unique point with every pair of numbers (cf. p. 4).

2. Without plotting the point A $(1, -3)$, tell what quadrant it lies in.

3. In which quadrants can a point lie if its abscissa is positive?

4. What are the signs of the coordinates of points lying in the second quadrant? In the third? In the fourth?

5. What are the coordinates in the plane of the point with coordinate -5 regarded as a point of the x-axis?

Answer. The point has abscissa -5 and ordinate 0.

The next three problems are a bit harder.

6. Plot the points A $(4, 1)$, B $(3, 5)$, C $(-1, 4)$ and D $(0, 0)$, thereby obtaining the vertices of a square. What is the side length of the square? What is its area?[2] Find the coordinates of the midpoints of the sides of the square. Can you *prove* that $ABCD$ is a square? Find four more points (give their coordinates) serving as the vertices of a square.

7. Plot a regular hexagon $ABCDEF$, choosing the point A as the origin and directing the x-axis from A to B, with the segment AB as the unit of length. Find the coordinates of each vertex of the hexagon. How many solutions does the problem have?

8. Given three points A $(0, 0)$, B (x_1, y_1) and D (x_2, y_2) in the plane, what coordinates must C have in order for the quadrilateral $ABCD$ to be a parallelogram?

5. Relations Between Coordinates

If both coordinates of a point are known, then the position of the point in the plane is completely determined. What can be said about the position of a point if only one of its coordinates is known? For example, how do we find all points with abscissas equal to 3 or all points one of whose coordinates equals 3?

---

2. As the unit of area, choose the area of a square whose side length equals the unit of length along the coordinate axes.

In general, specifying one of the two coordinates determines a curve. This fact is basic to the plot of Jules Verne's novel "Captain Grant's Children."[3] The heroes of the book only know one coordinate (the latitude) of the point where a shipwreck occurred, and hence to inspect all possible locations of the ship they are compelled to circumnavigate the globe along the circle of latitude 37° 11′.

Most often relations between coordinates determine a whole set of points rather than a single point. For example, if we plot all points whose abscissas equal their ordinates, i.e., points whose coordinates satisfy the equation

$$x = y,$$

we obtain the straight line bisecting the first and third quadrants (see Fig. 8).

Fig. 8

Sometimes we say the "locus of all points" instead of the "set of all points." For example, as just noted, the locus of all points whose coordinates satisfy the relation

$$x = y$$

is the bisector of the first and third quadrants.

It should not be thought that every relation between coordinates determines a curve in the plane. For example, you can easily convince yourself that the relation

$$x^2 + y^2 = 0$$

determines a unique point, namely the origin of coordinates. On the other hand, there is no point satisfying the relation

$$x^2 + y^2 = -1,$$

which determines the so-called "empty" set.

---
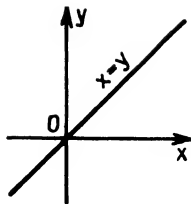
3. "Les Enfants du Capitaine Grant; Voyage Autour du Monde", Paris (1868). English translation published by Lippincott, Philadelphia (1874).

$x^2 - y^2 = 0$

(a)

$x^2-y^2>0$ $x^2-y^2>0$

(b)
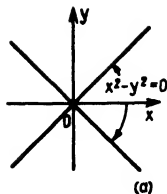
Fig. 9

The relation

$$x^2 - y^2 = 0$$

specifies a pair of orthogonal lines in the plane (see Fig. 9 a), while the relation

$$x^2 - y^2 > 0$$

determines a whole region (see Fig. 9 b).
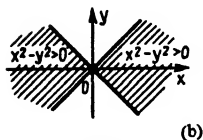
PROBLEMS

1. What sets of points are determined by the relations[4]

a) $|x| = |y|$; b) $\dfrac{x}{|x|} = \dfrac{y}{|y|}$; c) $|x| + x = |y| + y$;

d) $[x] = [y]$; e) $x - [x] = y - [y]$; f) $x - [x] > y - [y]$?

Answer. f) The shaded area shown in Figure 10.
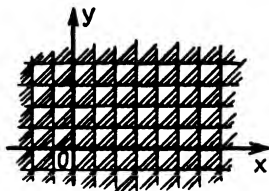
Fig. 10

2. A straight road separates a meadow from a marsh. Suppose a man can walk 3 miles an hour along the road, 2 miles an hour along the meadow and 1 mile an hour along the swamp. If the man starts walking from a point on the road, find the region consisting of all points which the man can reach within 1 hour.

---

4. The symbol $[x]$ denotes the *integral part* of x, i.e., the largest integer not exceeding x. For example, $[3.5] = 3$, $[5] = 5$, $[-2.5] = -3$.

3. The plane is divided by the coordinate axes into four quadrants. Suppose it is possible to move with velocity a in the first and third quadrants (including the coordinate axes) and with velocity b in the second and fourth quadrants (excluding the coordinate axes). Find the set of points that can be reached from the origin in a given time if

a) The velocity a is twice as large as b;

b) The velocities satisfy the formula

$$a = \sqrt{2}b.$$

6. Distance in Two Dimensions

We can now talk about points in numerical language. For example, there is no need to say "Take the point lying 3 units to the right of the y-axis and 5 units below the x-axis." It is enough to say simply "Take the point $(3, -5)$." As already noted, this has definite advantages. For example, a plane figure can now be transmitted by telegraph and then fed into a computing machine, which knows nothing at all about figures but understands numbers very well.

In the preceding section, we used relations between numbers to specify various sets of points in the plane. Continuing in this vein, we now translate a number of other geometric concepts into numerical language.

First we consider the simple problem of finding the distance between two points in the plane. As always, we assume that the points are specified by their coordinates, so that the problem is to find a rule expressing the distance between two points in terms of their coordinates. To find this rule, of course, we are allowed to draw a figure, but the rule itself must have nothing to do with figures. Instead it can only indicate which operations (and in what order) are to be performed on the given numbers, the coordinates of the points, in order to find the distance between the points.

Some readers may find this way of solving the problem strange and artificial. They will say "The points are specified by coordinates, but so what? Just plot the points, and then take a ruler and measure the distance between them!" Admittedly, this is sometimes a pretty good method However, again suppose we are dealing with a computing machine. It contains no ruler and it doesn't make drawings, but this is no handicap at all since it can calculate so rapidly (modern machines can perform tens of thousands of arithmetic operations per second). Our way of posing the problem of finding the distance between two points sees to it that the rule involves nothing but instructions which can be carried out by the machine.

First we solve the special case where one of the two points lies at the origin. Then the distance $\varrho\,(O, A)$ between the origin $O\,(0, 0)$ and the point $A(x, y)$ is the hypotenuse of a right triangle with legs of length x and y. Therefore, by the Pythagorean theorem,

$$\varrho\,(O, A) = \sqrt{x^2 + y^2}.$$

The rule expressed by this formula obviously satisfies all our requirements and in particular is suitable for use on computing machines, which are capable of multiplying numbers, forming sums and taking square roots.

Next we turn to the general case:

Problem. Find the distance $\varrho\,(A, B)$ between the two points $A\,(x_1, y_1)$ and $B\,(x_2, y_2)$.

Solution. Let A_1, B_1, A_2, B_2 be the projections of the points A and B onto the coordinate axes (see Fig. 11), and let C denote the point of intersection of the lines AA_1 and BB_2. Applying the Pythagorean theorem to the right triangle ABC, we obtain[5]

$$\varrho^2\,(A, B) = \varrho^2\,(A, C) + \varrho^2\,(B, C). \tag{1}$$

Fig. 11

---

5. By $\varrho^2\,(A, B)$ is meant the square of the distance $\varrho\,(A, B)$, and similarly for $\varrho^2\,(A, C)$ and $\varrho^2\,(B, C)$.
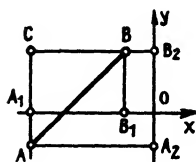
Moreover $\varrho(A, C)$, the length of the segment AC, equals the length of the segment $A_2 B_2$. But the points A_2 and B_2 lie on the y-axis and have coordinates y_1 and y_2, respectively. Therefore, by the formula derived in Sec.3, the distance between A_2 and B_2 equals $|y_1 - y_2|$. A similar argument shows that $\varrho(B, C)$, the length of the segment BC, equals $|x_1 - x_2|$. Substituting these values of $\varrho(A, C)$ and $\varrho(B, C)$ into (1), we obtain

$$\varrho^2(A, B) = (x_1 - x_2)^2 + (y_1 - y_2)^2.$$

Therefore $\varrho(A, B)$, the distance between the points $A(x_1, y_1)$ and $B(x_2, y_2)$, is given by the formula

$$\varrho(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \qquad (2)$$

Remark 1. The above considerations apply equally well to any other position of the points A and B, and not just that shown in Figure 11. You should draw another figure (for example, choosing the point A in the first quadrant and the point B in the second quadrant) and verify that the same argument carries over word for word without changing any of the notation.

Remark 2. The formula on p. 8 for the distance between two points on the coordinate line can be written in the form[6]

$$\varrho(A, B) = \sqrt{(x_1 - x_2)^2},$$

analogous to (2).

---

6. Here we use the fact that

$$\sqrt{x^2} = |x|,$$

where $\sqrt{x^2}$ denotes the *positive* square root. Failure to observe this convention by writing $\sqrt{x^2} = x$ instead can lead to false conclusions. For example, the following sequence of implications (read the symbol $\Rightarrow$ as "implies") contains an error of this kind (find it!):

$$1 - 3 = 4 - 6 \Rightarrow 1 - 3 + \tfrac{9}{4} = 4 - 6 + \tfrac{9}{4}$$
$$\Rightarrow (1 - \tfrac{3}{2})^2 = (2 - \tfrac{3}{2})^2 \Rightarrow \sqrt{(1 - \tfrac{3}{2})^2} = \sqrt{(2 - \tfrac{3}{2})^2}$$
$$\Rightarrow 1 - \tfrac{3}{2} = 2 - \tfrac{3}{2} \Rightarrow 1 = 2.$$

PROBLEMS

1. Find the distance between each of the points $(12, 5)$, $(-3, 15)$, $(-4, -7)$ and the origin.

2. Prove that the points $A(3, 6)$, $B(-2, 4)$ and $C(1, -2)$ are collinear.

Hint. Show that one of the sides of the "triangle" ABC is equal to the sum of the other two sides.

3. Use the formula for the distance between two points to prove the familiar theorem that the sum of the squares of the sides of a parallelogram equals the sum of the squares of the diagonals.

Hint. Choose one of the vertices of the parallelogram as the origin of coordinates and use the result of Problem 8, p. 12. Then the proof of the theorem reduces to verification of a simple algebraic identity. Which one?

4. Use the coordinate method to prove the following theorem: If $ABCD$ is a rectangle, then the formula $(AM)^2 + (CM)^2 = (BM)^2 + (DM)^2$ holds for any point M. What is the most convenient choice of axes?

7. Plane Figures

In Section 5, we gave several examples of how relations between coordinates define plane figures. We now pursue this study further.

Every figure will be regarded as the totality of points belonging to some set, and specifying the figure will mean giving a rule telling whether or not a given point belongs to this set. For example, to find such a rule for the case of a circle, we use the definition of a circle as a set of points all at the same distance R (the radius of the circle) from some point C (the center of the circle). In other words, a necessary and sufficient condition for a point $M(x, y)$ to lie on the circle with center $C(a, b)$
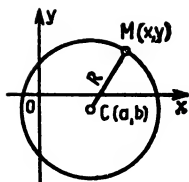
Fig. 12

and radius R is that $\varrho\,(M, C)$ equal R (see Fig. 12). But according to formula (1), p. 16, this condition can be written as

$$\sqrt{(x - a)^2 + (y - b)^2} = R,$$

or equivalently as

$$(x - a)^2 + (y - b)^2 = R^2. \tag{3}$$

Thus to determine whether a given point lies on the circle in question, we need only determine whether the coordinates of the point satisfy equation (3). If so, the point lies on the circle, but otherwise it does not. In other words, by using (3) we can determine whether or not any given point belongs to the circle. For this reason, (3) is called the *equation of the circle with center C (a, b) and radius R.*

PROBLEMS

1. Write the equation of the circle with center $C(-2, 3)$ and radius 5. Does this circle pass through the point $(2, -1)$?

2. Show that

$$x^2 + 2x + y^2 = 0$$

is the equation of a circle in the plane. Find the center and radius of the circle.

   Hint. Write the equation in the form

$$(x^2 + 2x + 1) + y^2 = 1$$

or

$$(x + 1)^2 + y^2 = 1.$$

3. What is the set of points determined by the relation

$$x^2 + y^2 \leqslant 4x + 4y? \tag{4}$$

   Answer. Writing (4) as

$$x^2 - 4x + 4 + y^2 - 4y + 4 \leqslant 8$$

or

$$(x - 2)^2 + (y - 2)^2 \leqslant 8, \tag{5}$$

we find that the distance between a point (x, y) of the given set and the point $(2, 2)$ is less than or equal to $\sqrt{8}$. Obviously the set of all such points is a disk of radius $\sqrt{8}$ with center $(2, 2)$. The boundary of the disk also belongs to the set, since equality is permitted in (5).

We have just seen how a circle can be specified by an equation of a certain kind. In the same way, other curves can be specified by appropriate equations. For example, as already noted on p. 14, the equation

$$x^2 - y^2 = 0 \tag{6}$$

specifies a pair of straight lines. Let's examine this in a bit more detail. If $x^2 - y^2 = 0$, then $x^2 = y^2$ and hence

$$|x| = |y|. \tag{7}$$

Conversely, if $|x| = |y|$ then $x^2 - y^2 = 0$. Therefore the relations (6) and (7) are equivalent. But the absolute value of the abscissa of a point is its distance from the y-axis, while the absolute value of the ordinate of a point is its distance from the x-axis. In other words, the points satisfying $|x| = |y|$ are equidistant from the two coordinate axes and hence must lie on the two bisectors of the angles formed by the coordinate axes. Conversely, it is clear that the coordinates of any point on either of the two bisectors satisfies the relation $x^2 = y^2$. Therefore (6) is called the equation of the two bisectors.

Some of you may be familiar with the fact that the equation

$$y = x^2 \tag{8}$$

is satisfied by the points of a parabola with its vertex at the origin, and only by such points. Therefore (8) is called the equation of a parabola.

In general, by the *equation of a curve* we mean an equation in the variables x and y which becomes an identity whenever x and y are replaced by the coordinates of any point of the curve,

but is not satisfied if x and y are replaced by the coordinates of a point which does not lie on the curve.

For example, even without knowing what curve is specified by the equation

$$(x^2 + y^2 + y)^2 = x^2 + y^2, \tag{9}$$

we can say that the curve passes through the origin, since (9) is satisfied by $x = 0$, $y = 0$. On the other hand, the point $(1, 1)$ does not lie on the curve since $(1^2 + 1^2 + 1)^2 \neq 1^2 + 1^2$. If you are interested in the appearance of the curve specified by (9), take a look at Figure 13. The curve is shaped like a heart and hence is called a *cardioid*.

Thus a computing machine in love might well print out the equation of a cardioid, or even the equations of the flower-shaped curves shown in Figure 14 (as a kind of "mathematical bouquet"!). The equations of these mathematical flowers will be given in Section 9, after we have become acquainted with coordinates of a different kind, called *polar coordinates*.

Fig. 13

(a)

(b)
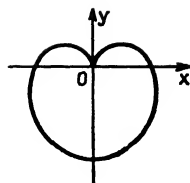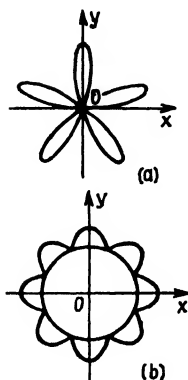
Fig. 14

8. Coordinates as a Tool for Solving Geometric Problems

By transcribing geometric problems into coordinate language, they can then be solved by using algebraic methods. After this transcription has been made, it turns out that most problems involving straight lines and circles lead to equations of the first and second degrees, which can be handled very simply and generally. (It should be noted that the art of solving algebraic equations was already highly developed in the seventeenth century, at the time of the discovery of the coordinate method. In fact, mathematicians had by then learned to solve algebraic equations of the third and forth degrees. Thus René Descartes, the French mathematician and philosopher who invented the coordinate method, was prompted to say "I have solved all problems," referring to the geometric problems of his day.)

The following example illustrates how a geometric problem can be made algebraic.

Problem 1. Find the center of the circle circumscribed about a given triangle ABC.

Solution. Choose the point A as the origin, and direct the axis of abscissas from A to B. Then B is the point $(c, 0)$, where c is the length of the segment AB. Let the point C have abscissa q and ordinate h, and let the required circle have radius R and center at the point (α, β). Then each of the points $A\ (0, 0)$, $B\ (c, 0)$ and $C\ (q, h)$ lies on the circle, i.e., the distance between each of these points and the center (α, β) of the circle equals the radius R. This is expressed algebraically by writing

$$\alpha^2 + \beta^2 = R^2,$$

$$(c - \alpha)^2 + \beta^2 = R^2, \tag{10}$$

$$(q - \alpha)^2 + (h - \beta)^2 = R^2.$$

The same conditions are easily obtained by writing the equation of the circle with center (α, β) and radius R, i.e.,

$$(x - \alpha)^2 + (y - \beta)^2 = R^2,$$

and then replacing x and y in turn by the coordinates of the points A, B and C lying on the circle. The solution of the system (10) of three equations in three unknowns is easily found to be

$$\alpha = \frac{c}{2}, \quad \beta = \frac{q^2 + h^2 - cq}{2h},$$

$$R = \frac{\sqrt{(q^2 + h^2)\,[(q - c)^2 + h^2]}}{2h}.$$

The problem is now solved (without resorting to a figure!), since we know the coordinates α and β of the center of the circle.

At the same time we have found a formula for the radius of the circle circumscribed about the triangle. This formula can be

simplified by noting that

$$\sqrt{q^2 + h^2} = \varrho(A, C), \quad \sqrt{(q - c)^2 + h^2} = \varrho(B, C),$$

while the number h is the altitude of the triangle ABC dropped from the vertex C. Denoting the lengths of the sides BC and AC of the triangle by a and b, we find that the radius R is given by the elegant and convenient formula

$$R = \frac{ab}{2h}. \tag{11}$$

Noting further that $hc = 2S$, where S is the area of the triangle, we can also write (11) in the form

$$R = \frac{abc}{4S}.$$

Next we consider a problem whose solution by purely geometric methods is quite complicated. However, interestingly enough, the problem is easily solved when translated into coordinate language.

Problem 2. Given two points A and B in the plane, find the locus of all points M which are twice as far from A as from B.

Solution. Choose a coordinate system in the plane such that the origin is at A and the positive x-axis goes from A to B. Choose the segment AB as the unit of length. Then A is the point $(0, 0)$, while B is the point $(1, 0)$. If the coordinates of the point M are denoted by x and y, we can write the condition

$$\varrho(A, M) = 2\varrho(B, M)$$

in the form

$$\sqrt{x^2 + y^2} = 2\sqrt{(x - 1)^2 + y^2}, \tag{12}$$

which is the equation of the required locus. To find the set described by (12), we transform it into a form familiar to you. Squaring both sides of (12), eliminating parentheses and combining similar terms, we find that (12) becomes

$$3x^2 - 8x + 4 + 3y^2 = 0,$$

which can be rewritten as

$$x^2 - \frac{8}{3}x + \frac{16}{9} + y^2 = \frac{4}{9}$$

or

$$\left(x - \frac{4}{3}\right)^2 + y^2 = \left(\frac{2}{3}\right)^2.$$

The last equation will be recognized at once as the equation of a circle of radius $\frac{2}{3}$, with its center at the point ($\frac{4}{3}$, 0), i.e., the required locus is a circle.

Remark 1. Let k be a positive number different from 1. Then the same method of proof shows that the locus of all points M which are k times as far from A as from B is a circle. The case $k = 1$ is excluded, since, as you know, the locus of all points M equidistant from A and B is a straight line (prove this analytically).

Remark 2. To appreciate the power of the coordinate method, try to solve Problem 2 geometrically.

Hint. From the point M draw the bisectors of the interior and exterior angles of the triangle AMB, and let K and L be the points of intersection of these bisectors with the line AB. Prove that the position of these points is independent of the choice of the point M on the required locus. Prove that the angle KML equals 90°.

Comment. The ancient Greeks knew how to cope with problems of this type. In fact, a geometric solution of Problem 2 can be found in the treatise "On Circles" by the Greek mathematician Apollonius (second century B.C.).

PROBLEM

Given two points A and B in the plane, find the locus of all points M such that

$$\varrho^2(A, M) - \varrho^2(B, M) = c.$$

For what values of c does a solution exist?

9. Other Coordinate Systems

Besides rectangular Cartesian coordinates, there are other coordinate systems in the plane. One such system, called *oblique Cartesian coordinates,* is shown in Figure 15, which makes it clear how to determine the coordinates x and y. There are also cases where it is necessary to choose different units of length along the coordinates axes.
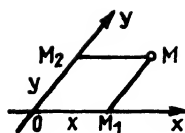
Fig. 15

Next we consider coordinates which differ more radically from Cartesian coordinates. Typical of such coordinates are *polar coordinates* (already mentioned on p. 21) which are defined as follows: Choose a coordinate line in the plane (see Fig. 16), calling the origin of coordinates (the point O) the *pole* and the line itself the *polar axis.* To specify the position of the point M, we need only give two numbers, the *radius vector ϱ* (the distance from the pole to M) and the *polar angle φ* (the angle between the polar axis and the ray OM measured in the counterclockwise direction.)[7] For example, in Figure 16 the radius vector equals 3.5 (e is the unit of length), while the polar angle equals 225° or $5\pi/4$. For brevity, the point with polar coordinates ϱ, φ is often called "the point (ϱ, φ)."
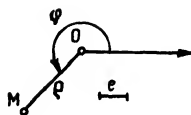
Fig. 16

Here we measure the polar angle φ in *radians* as well as in degrees. In radian measure the unit of angular measurement is taken to be the *radian,* the central angle subtended by a circular arc of length 1 (the radius of the circle is assumed to be 1). Thus the angle of 360° corresponding to the whole circle (of radius 1) has radian measure 2π, the angle 180° has radian measure π, a right angle has radian measure $\pi/2$, the angle 45° has radian measure $\pi/4$, and so on. Clearly a radian equals

$$\frac{180°}{\pi} \approx \frac{180°}{3.14} \approx 57° \; 17' \; 45''.$$

In many problems it turns out to be much more convenient to measure angles in radians than in degrees.

---

7. The Greek letter φ is read "phi."

clockwise direction) and at the same time moves away from the pole. The equation

$$\varrho = \frac{1}{\varphi}$$

represents another kind of spiral, shown in Figure 17b. Here ϱ is large if φ is near zero, while ϱ decreases as φ increases and is small if φ is large. Therefore the spiral "winds around" the point O as φ gets larger and larger.
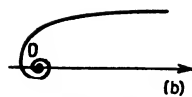

(b)

Fig. 17b

In some cases, polar coordinates are more convenient than Cartesian coordinates, provided you know a little trigonometry. For example, in polar coordinates the cardioid shown in Figure 13, p. 21 has the equation

$$\varrho = 1 - \sin \varphi,$$

which is considerably simpler than its equation in Cartesian coordinates [equation (9), p. 21]. As for the pretty flowers shown in Figure 14, p. 21, in polar coordinates they have the simple equations

$$\varrho = \sin 5\varphi \quad (\text{Fig. 14a}),$$

$$(\varrho - 2)(\varrho - 2 - |\cos 3\varphi|) = 0 \quad (\text{Fig. 14b}).$$

PROBLEM

What are the sets specified by the relations

a) $\varrho = \sin \varphi$;   b) $\varrho (\cos \varphi + \sin \varphi) + 1 = 0$?

Unlike the case of Cartesian coordinates, polar coordinates do not establish an one-to-one correspondence between the points of the plane and pairs of numbers ϱ and φ, written in a definite order (first ϱ, then φ). In fact, adding any integral multiple of 2π (i.e., any integral multiple of 360°) to the angle φ obviously has no effect on the direction of the ray joining the pole to a given point (ϱ, φ). In other words, given any number $\varrho > 0$ and any integer k, the point with coordinates ϱ, φ is the same as the point with coordinates $\varrho, \varphi + 2k\pi$.

In this connection, we give another example of a situation where the correspondence between points and their coordinates fails to be unique.

clockwise direction) and at the same time moves away from the pole. The equation

$$\varrho = \frac{1}{\varphi}$$

represents another kind of spiral, shown in Figure 17b. Here ϱ is large if φ is near zero, while ϱ decreases as φ increases and is small if φ is large. Therefore the spiral "winds around" the point O as φ gets larger and larger.
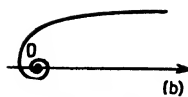

(b)

Fig. 17b

In some cases, polar coordinates are more convenient than Cartesian coordinates, provided you know a little trigonometry. For example, in polar coordinates the cardioid shown in Figure 13, p. 21 has the equation

$$\varrho = 1 - \sin \varphi,$$

which is considerably simpler than its equation in Cartesian coordinates [equation (9), p. 21]. As for the pretty flowers shown in Figure 14, p. 21, in polar coordinates they have the simple equations

$$\varrho = \sin 5\varphi \quad (\text{Fig. 14a}),$$

$$(\varrho - 2)(\varrho - 2 - |\cos 3\varphi|) = 0 \quad (\text{Fig. 14b}).$$

PROBLEM

What are the sets specified by the relations

a) $\varrho = \sin \varphi$; b) $\varrho (\cos \varphi + \sin \varphi) + 1 = 0$?

Unlike the case of Cartesian coordinates, polar coordinates do not establish an one-to-one correspondence between the points of the plane and pairs of numbers ϱ and φ, written in a definite order (first ϱ, then φ). In fact, adding any integral multiple of 2π (i.e., any integral multiple of 360°) to the angle φ obviously has no effect on the direction of the ray joining the pole to a given point (ϱ, φ). In other words, given any number $\varrho > 0$ and any integer k, the point with coordinates ϱ, φ is the same as the point with coordinates $\varrho, \varphi + 2k\pi$.

In this connection, we give another example of a situation where the correspondence between points and their coordinates fails to be unique.
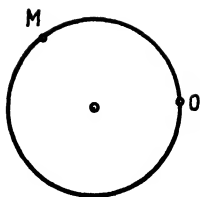
Fig. 18

In the Introduction we already talked about defining coordinates on curves, and in Chapter 1 we showed how coordinates are defined on the simplest curve of all, the straight line. We now devise a way of introducing coordinates on another simple curve, namely the circle. Just as in Sec. 1, we begin by choosing a point on the circle as the origin of coordinates (the point O in Figure 18). As usual, the counterclockwise direction will be regarded as the positive direction of motion along the circle. For the unit of length along the circle, we make the natural choice, i.e., the radius of the circle. Then the coordinate of a point M on the circle will be the length of the arc OM, taken with a plus sign if the rotation carrying O into M is positive and with a minus sign otherwise.

An important difference between these coordinates and the coordinates of points on a straight line is now immediately apparent: *The correspondence between points and numbers (coordinates) is not one-to-one.* Clearly there is a unique point of the circle corresponding to each number. However, the converse is not true, i.e., there are infinitely many numbers corresponding to the same point of the circle. In fact, given any number a, the point of the circle with coordinate a is found by starting from the origin and marking off a distance of a units along the circle in the positive or negative direction, depending on whether a is positive or negative. Let M be the point found in this way. Then clearly we get the same point M if we measure off a distance of $\pm 2\pi$ starting from M, since this corresponds to making a complete circuit around the circle (whose circumference equals 2π). By the same token, there are infinitely many numbers corresponding to the same point M with coordinate a, namely the numbers

$$\ldots, a - 4\pi, a - 2\pi, a, a + 2\pi, a + 4\pi, \ldots$$

obtained by adding integral multiples of 2π to a.

The coordinate a coincides with the angle φ of a polar coordinate system, provided φ is measured in radians. Thus this example sheds further light on why a point does not have unique polar coordinates.

Coordinates in Space

10. Coordinate Axes and Planes

Two coordinate axes are enough to specify the position of a point in the plane, but to specify the position of a point in space we now need *three* coordinate axes. These three axes, called the *x-axis*, the *y-axis* and the *z-axis*,[1] all pass through the same point *O*, called the *origin of coordinates*. Moreover the three axes are mutually perpendicular, i.e., each axis is perpendicular to the other two, and the directions of the axes are usually chosen in such a way that the positive *x*-axis is carried into the positive *y*-axis by a counterclockwise rotation of 90° as seen from the positive *z*-axis (see Fig. 19).

In space it is convenient to consider *coordinate planes* as well as coordinate axes. There are three such planes, shown in Figure 20, each passing through a pair of coordinate axes:

1) The *xy-plane*, passing through the *x* and *y*-axes and consisting of all points of the form $(x, y, 0)$ where x and y are arbitrary numbers;

2) The *xz-plane*, passing through the *x* and *z*-axes and consisting of all points of the form $(x, 0, z)$ where x and z are arbitrary numbers;

3) The *yz-plane*, passing through the *y* and *z*-axes and
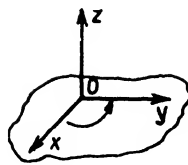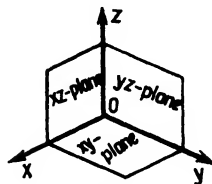

Fig. 19

Fig. 20

---

1. As in the case of the plane, the *x*-axis is called the *axis of abscissas* and the *y*-axis is called the *axis of ordinates*. The *z*-axis has no commonly accepted name, but we suggest the term *axis of applicates*.

consisting of all points of the form $(0, y, z)$ where y and z are arbitrary numbers.
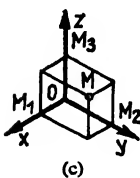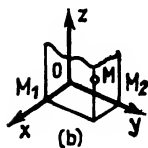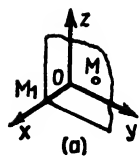
Given any point M in space, we can now find three numbers x, y and z which serve as its coordinates. To find the first coordinate, we draw the plane through M parallel to the yz-plane (and at the same time perpendicular to the x-axis). Let M_1 be the point in which this plane intersects the the x-axis (see Fig. 21 a). Then M_1 is a point of the axis Ox, and as such has a definite coordinate x called the *abscissa* of M. Similarly, to find the second coordinate, we draw the plane through M parallel to the xz-plane (and at the same time perpendicular to the y-axis). Let M_2 be the point in which this plane intersects the y-axis (see Fig. 21 b). Then M_2 is a point of the axis Oy, and as such has a definite coordinate y called the *ordinate* of M. Finally, we draw the plane through M parallel to the xy-plane (and perpendicular to the z-axis). Suppose this plane intersects the z-axis in the point M_3 (see Fig. 21 c). Then M_3 is a point of the axis Oz, and as such has a definite coordinate z which we call the *applicate* of M (pursuing the suggestion made in footnote 1).

Thus to every point M in space there corresponds a triple of numbers x, y and z, called the *rectangular Cartesian coordinates* of M. Conversely, given any triple of numbers x, y and z, there is a point M in space with x as abscissa, y as ordinate and z as applicate. To find M, we merely reverse the steps in the construction shown in Figure 21, first plotting the points M_1, M_2 and M_3 with coordinates x, y and z along the axes Ox, Oy and Oz. Next through the points M_1, M_2 and M_3, we draw planes parallel to the yz, xz and xy-planes. Then the point in which three planes intersect is clearly the point M with x, y and z as its coordinates.

This establishes an one-to-one correspondence (as defined on p. 4) between the points of space and triples of numbers x, y and z, written in a definite order (first x, then y and finally z).

Fig. 21

The subject of coordinates in space requires some background in solid geometry, and hence is a bit harder than that of coordinates in the plane. However, everything about geometry needed for our purposes is simple, intuitive and readily intelligible. Those of you who feel the need for more rigor can consult your high school course on solid geometry, where, for example, it is shown that the points M_1, M_2 and M_3 in which the coordinate axes intersect the planes drawn through M parallel to the coordinate planes are just the projections of the point M onto the coordinate axes, i.e., the feet of the perpendiculars dropped from M onto the coordinate axes. Thus the coordinates of a point M in space can be defined in just the same way as for coordinates in the plane, namely as the coordinates of the projections of M onto the axes Ox, Oy and Oz, regarded as points of these axes.

Many of the formulas derived in Chapter 2 for the case of the plane carry over to the case of space with only slight modifications. For example, the distance between the two points $A\,(x_1, y_1, z_1)$ and $B\,(x_2, y_2, z_2)$ is given by the formula

$$\varrho\,(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

The derivation of this formula closely resembles that of the analogous formula for the plane (see pp. 16-17), and is left as an exercise.

In particular, the distance between the origin $O\,(0, 0, 0)$ and the point $A\,(x, y, z)$ is given by

$$\varrho\,(O, A) = \sqrt{x^2 + y^2 + z^2}.$$

PROBLEMS

1. Which of the eight points

$$(1, 1, 1), \quad (1, 1, -1), \quad (1, -1, 1), \quad (1, -1, -1),$$
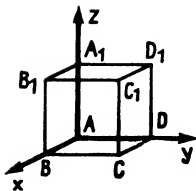$$(-1, 1, 1), (-1, 1, -1), (-1, -1, 1), (-1, -1, -1)$$

Fig. 22

is furthest from the point $(1, 1, 1)$ and what is its distance from $(1, 1, 1)$? Apart from $(1, 1, 1)$, which points are closest to $(1, 1, 1)$, and what is the distance between these points and $(1, 1, 1)$?

2. Draw a cube, directing the coordinate axes along the three edges emanating from some vertex. Label the vertices of the cube by the letters $A, B, C, D, A_1, B_1, C_1$ and D_1, as shown in Figure 22. With the edge of the cube as the unit of length, find the coordinates of

a) All the vertices of the cube;

b) The midpoint of the edge CC_1;

c) The point of intersection of the diagonals of the face AA_1B_1B.

3. What is the distance between the vertex $(0, 0, 0)$ of the cube in Problem 2 and the point of intersection of the diagonals of the face BB_1C_1C?

4. Which of the points

$$(1, 0, 5), \ (3, 0, 1), \ (\tfrac{1}{3}, \tfrac{3}{4}, \tfrac{2}{5}), \ (\tfrac{7}{5}, \tfrac{1}{2}, \tfrac{3}{2}), \ (\tfrac{2}{5}, -\tfrac{1}{2}, 0), \ (1, \tfrac{1}{2}, \tfrac{1}{3})$$

lie inside the cube in Problem 2 or on its boundary? Which lie outside the cube?

5. Write the relations satisfied by the coordinates x, y, z of the points lying inside the cube in Problem 2 or on its boundary.

Answer. The coordinates x, y, z take all values between 0 and 1 inclusive, i.e., they satisfy the inequalities

$$0 \leqslant x \leqslant 1, \quad 0 \leqslant y \leqslant 1, \quad 0 \leqslant z \leqslant 1.$$

11. Space Figures

With the aid of coordinates in space (just as in the plane), we can use numbers and numerical relations to specify not only points, but also curves, surfaces and other sets of points. For example, what is the set of points obtained if two coordinates are specified and the other is left arbitrary? The conditions

$$x = a, \quad y = b,$$

where a and b are fixed numbers (say $a = 5$, $b = 4$) determine a straight line in space parallel to the z-axis (see Fig. 23). In fact, every point of such a line has the same abscissa a and the same ordinate b, while the z-coordinate takes arbitrary values. Similarly, the conditions

$$y = b, \quad z = c$$

determine a straight line parallel to the x-axis, and the conditions

$$z = c, \quad x = a$$

Fig. 23

determine a straight line parallel to the y-axis.

In the same vein, what is the set of points obtained if we specify only one coordinate, say $z = 1$? As is clear from Figure 24, the answer is a plane parallel to the xy-plane (i.e., the plane passing through the x and y-axes), at distance 1 from the xy-plane in the direction of the positive z-axis.

Fig. 24

We now analyze a number of other examples showing how various space figures can be specified by using equations and other relations between coordinates.

1. Interpret the equation

$$x^2 + y^2 + z^2 = R^2. \tag{1}$$

Since the distance from the origin of coordinates to the point (x, y, z) is given by $\sqrt{x^2 + y^2 + z^2}$, the geometric interpretation of (1) is clear, i.e., the distance between the origin and any point (x, y, z) satisfying (1) equals R. In other words, the set of all points satisfying (1) is a sphere of radius 1 with its center at the origin.

2. Find the set of points with coordinates satisfying the inequality

$$x^2 + y^2 + z^2 < 1. \tag{2}$$

Since (2) means that the distance from the origin to the point (x, y, z) is less than 1, the set in question consists of all points lying inside a sphere of unit radius with its center at the origin.
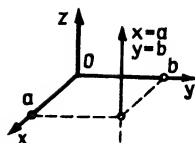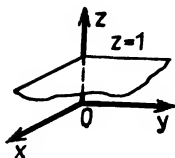
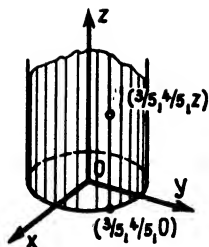3. What is the set of points determined by the equation

$$x^2 + y^2 = 1? \tag{3}$$

We begin by considering only points of the xy-plane (i.e., points such that $z = 0$) satisfying (3). Then as we know from p. 19, (3) determines a circle of radius 1 with its center at the origin. The x and y-coordinates of every point of this circle satisfy (3), while the z-coordinate vanishes. For example, the point $(\frac{3}{5}, \frac{4}{5}, 0)$ satisfies (3), as shown in Figure 25, and this can be used at once to find many other points satisfying the same equation. In fact, since equation (3) does not contain z, the points $(\frac{3}{5}, \frac{4}{5}, 10)$ and $(\frac{3}{5}, \frac{4}{5}, -5)$ also satisfy (3), and more generally so do all points $(\frac{3}{5}, \frac{4}{5}, z)$, where the value of z is completely arbitrary. All these points lie on the line passing through the point $(\frac{3}{5}, \frac{4}{5}, 0)$ parallel to the z-axis.

In just the same way, starting from any point $(x^*, y^*, 0)$ of our circle lying in the xy-plane, we can find many other points satisfying (3). In fact, we need only draw a straight line through $(x^*, y^*, 0)$ parallel to the z-axis. Every point of this straight line has the same values of x and y as the point $(x^*, y^*, 0)$ of the circle, while the value of z is arbitrary, i.e., every point of the straight line is of the form (x^*, y^*, z). But since (3) does not involve z, while the point $(x^*, y^*, 0)$ satisfies (3), the point (x^*, y^*, z) must also satisfy (3). Moreover, it is clear that every point satisfying (3) can be found in this way.

Thus, finally, the set of points satisfying (3) is obtained as follows: In the xy-plane we take the circle of radius 1 with its center at the origin. Then through every point of the circle we draw a straight line parallel to the z-axis, thereby obtaining a *cylindrical surface* (see Fig. 25).

4. We have seen that in three dimensions a single equation usually determines a surface. However, there are exceptions to this rule. For example, the equation $x^2 + y^2 = 0$ is only satisfied by the points of a curve, namely the z-axis, since $x^2 + y^2 = 0$ implies that $x = y = 0$, and every point with zero abscissa and

Fig. 25

ordinate lies on the z-axis (conversely, every point of the z-axis has zero abscissa and ordinate). The equation $x^2 + y^2 + z^2 = 0$ represents a point (the origin of coordinates), while the equation $x^2 + y^2 + z^2 = -1$ determines the empty set.

5. What happens if we consider points whose coordinates satisfy a system of equations rather than a single equation? For example, consider the system

$$x^2 + y^2 + z^2 = 4,$$

$$z = 1. \tag{4}$$

The points satisfying the first of the equations (4) make up a sphere of radius 2 with its center at the origin, while the points satisfying the second of the equations (4) make up a plane parallel to the xy-plane at distance 1 from the xy-plane in the direction of the positive z-axis. Therefore the points satisfying both equations must lie on both the sphere $x^2 + y^2 + z^2 = 4$ and the plane $z = 1$, i.e., they must belong to the intersection of the sphere and the plane. In other words, the system (4) determines the curve in which the sphere and plane intersect, namely a circle (see Fig. 26).
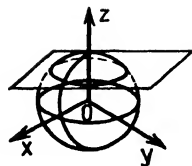
Fig. 26

6. How do we specify (in space) the circle of radius 1 which lies in the xz-plane and has its center at the origin?

As we have seen, the equation $x^2 + z^2 = 1$ determines a cylindrical surface. To confine ourselves to the points of the required circle, we have to add the extra condition $y = 0$, which from all the points of the cylinder selects out just those lying in the xz-plane (see Fig. 27). This gives the system

Fig. 27

$$x^2 + z^2 = 1,$$

$$y = 0.$$
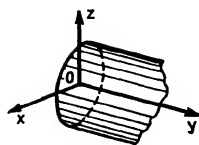
PROBLEMS

1. Which of the points

$$A\left(\sqrt{2}, \sqrt{2}, 0\right), \quad B\left(\sqrt{2}, \sqrt{2}, 1\right), \quad C\left(\sqrt{2}, \sqrt{2}, \sqrt{2}\right),$$
$$D\left(1, \sqrt{3}, 0\right), \quad E\left(0, \sqrt{3}, 1\right), \quad F\left(-1, -\sqrt{2}, 1\right)$$

belong to the sphere $x^2 + y^2 + z^2 = 4$? Which belong to the plane $z = 1$? Which belong to the curve in which the sphere and plane intersect?

2. Find the set of points in space determined by each of the following equations:

a) $z^2 = 1$; b) $y^2 + z^2 = 1$; c) $x^2 + y^2 + z^2 = 1$.

3. Do the following systems of equations all determine the same curve:

a) $x^2 + y^2 + z^2 = 1$, $y^2 + z^2 = 1$;

b) $x^2 + y^2 + z^2 = 1$, $x = 0$;

c) $y^2 + z^2 = 1$, $x = 0$?

4. How do we specify (in space) the bisector of the angle xOy? What is the set specified (in space) by the single equation $x = y$?

PART 2

SOME MODERN TOPICS

Exploration of the Fourth Dimension

12. Some General Remarks

Algebra and geometry, usually regarded by schoolboys as completely different subjects, are actually very intimately related. In fact, by using the coordinate method, we could give an entire course in high school geometry without drawing a single figure, resorting only to numbers and algebraic operations. A course in plane geometry would then begin with the words "By a point is meant a pair of numbers (x, y)..." We could then define a circle as a set of points satisfying an equation of the form

$$(x - a)^2 + (y - b)^2 = R^2,$$

a straight line as a set of points satisfying an equation of the form

$$ax + by + c = 0,$$

and so on. As a result, the theorems of plane geometry would all reduce to certain algebraic relations.

The discovery of the connection between algebra and geometry led to a veritable revolution in mathematics, revealing that mathematics is a unified science with no "Chinese walls" between its various branches. The invention of the coordinate method is attributed to the French philosopher and mathematician René Descartes (1596–1650). In the last part of a philosophical treatise published in 1637, Descartes describes

the coordinate method and its application to the solution of geometric problems. Descartes' ideas evolved into a special branch of mathematics now known as "analytic geometry," a designation which contains the basic idea of the subject. Thus analytic geometry is the branch of mathematics which uses analytic (i.e., algebraic) methods to solve geometric problems. Although analytic geometry is today a fully developed subject in definitive form, its key ideas continue to generate new branches of mathematics, like algebraic geometry which studies the properties of curves and surfaces specified by algebraic equations. Algebraic geometry is a flourishing subject which is far from achieving its final form. In fact, some basic new results discovered quite recently in algebraic geometry have had a great influence on other branches of mathematics.

13. Geometry as a Tool for Counting

In using analytic geometry to solve geometric problems, we concentrate on one aspect of the coordinate method, namely the analytic interpretation of geometric concepts made by translating geometric objects and relationships into numerical language. However, the converse aspect of the coordinate method, namely the geometric interpretation of numbers and numerical relationships, turns out to be no less important. In fact, the famous mathematician Hermann Minkowski (1864-1909) used geometric methods to find *integral solutions* of algebraic equations, i.e., solutions in which the unknowns are only allowed to be whole numbers. Much to the surprise of the mathematicians of his time, certain problems of number theory, previously regarded as very difficult, turned out to be simple and straightforward when approached from a geometric point of view.

To show how geometry can be used to solve algebraic

problems, we now pose the following simple question: *What is the number of integral solutions of the inequality*

$$x^2 + y^2 \leqq n?$$ (1)

For small values of n, the answer is easy. For example, for $n = 0$ there is only one solution $x = 0, y = 0$. For $n = 1$ there are four more solutions, namely

$$x = 0, y = 1; \quad x = 1, y = 0;$$

$$x = 0, y = -1; \quad x = -1, y = 0.$$

In other words, for $n = 1$ there are five solutions in all. For $n = 2$ there are four more solutions besides those already given:

$$x = 1, y = 1; \quad x = -1, y = 1;$$

$$x = 1, y = -1; \quad x = -1, y = -1.$$

Thus for $n = 2$ there are nine solutions in all. Continuing in this way, we arrive at the following table:

Number n	Number of solutions N	Ratio N/n
0	1	—
1	5	5
2	9	4.5
3	9	3
4	13	3.25
5	21	4.2
10	37	3.7
20	69	3.45
50	161	3.22
100	317	3.17

Consulting the table, we see that the number of solutions N increases with increasing n, but it would be difficult to guess
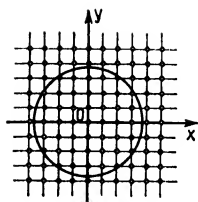
exactly how N depends on n. However, a glance at the third column of the table shows that it is reasonable to assume that the ratio N/n approaches a certain number as n increases. By interpreting the problem geometrically, we will now show that this assumption is actually justified and that the number approached by the ratio N/n is none other than the familiar constant $\pi = 3.14159265\ldots$! This is done as follows: If the pair of numbers (x, y) is regarded as a point in the plane (with abscissa x and ordinate y), then the inequality $x^2 + y^2 \leqq n$ means that the point (x, y) lies inside the disk K_n of radius $\sqrt{n}$ with its center at the origin (see Fig. 28). Thus our inequality has as many integral solutions as there are points with integral coordinates in the disk K_n. It is geometrically obvious that the points with integral coordinates are "uniformly distributed over the plane" with one point per unit area of the plane. Therefore it is clear that the number of solutions must be approximately equal to the area of the disk. This suggests the approximate formula

Fig. 28

$$N \approx \pi n. \tag{2}$$

A brief proof of formula (2) goes as follows: We divide the plane into unit squares by drawing lines parallel to the coordinate axes. Then the vertices of the squares are the points with integral coordinates. Suppose K_n contains N points with integral coordinates. With each of these points we associate the unit square which has the point as its upper right hand vertex. Let A_n be the figure formed by these squares (A_n is shaded in Fig. 29). Then the area of A_n obviously equals N, since A_n is made up of precisely N squares.
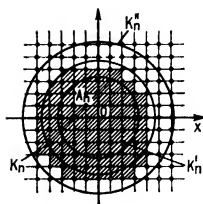
Fig. 29

Next we compare the area of A_n with that of the disk K_n. Besides K_n we consider two other disks with their centers at the origin, namely the disk K_n' of radius $\sqrt{n} - \sqrt{2}$ and the disk K_n'' of radius $\sqrt{n} + \sqrt{2}$. Then the figure A_n lies entirely inside the disk K_n'', while the disk K_n' lies entirely within A_n (prove this, using the fact that the diagonal of the unit square is of length $\sqrt{2}$). Therefore the area of A_n exceeds the area of K_n' and is less than the area of K_n'', i.e.,

$$\pi \left(\sqrt{n} - \sqrt{2}\right)^2 < N < \pi \left(\sqrt{n} + \sqrt{2}\right)^2.$$

This proves the approximate formula (2), and at the same time establishes the following estimate of the error in (2):

$$|N - \pi n| < 2\pi (\sqrt{2n} + 1).$$

Suppose we now rephrase our problem for the case of three unknowns: *What is the number of integral solutions of the inequality*

$$x^2 + y^2 + z^2 \leqslant n?$$ (3)

Again the answer is obtained very quickly by using the geometric interpretation, i.e., the number of solutions is approximately equal to $\tfrac{4}{3} \pi n \sqrt{n}$, the volume of a ball (solid sphere) of radius $\sqrt{n}$. It would be hard to obtain this result by purely algebraic means.

14. The Need for Four Dimensions

We now generalize the questions in the preceding section, and ask for the number of integral solutions of the inequality

$$x^2 + y^2 + z^2 + u^2 \leqslant n$$ (4)

in four unknowns. The same problem for two and three unknowns was solved by using a geometric interpretation. In fact, any solution of the inequality (1) in two unknowns is a pair of numbers (x, y) which we regarded as a point in the plane, while any solution of the inequality (3) in three unknowns is a triple of numbers (x, y, z) which we regarded as a point in space. Therefore shouldn't we continue to use this method? If so, a quadruple of numbers (x, y, z, u) ought to be regarded as a point in a "four-dimensional space." Then the inequality (4) can be interpreted as the condition for the points (x, y, z, u) to lie inside a "four-dimensional ball" of radius $\sqrt{n}$ with its center at the origin. As before, this reduces to a calculation of the volume of the ball (after dividing four-dimensional space into

"four-dimensional cubes"). By a method beyond the scope of this primer,[1] it can be shown that the volume of a four-dimensional ball of radius is $\frac{1}{2} \pi^2 R^4$. Therefore the number of integral solutions of (4) is just $\frac{1}{2} \pi^2 n^2$.

The above considerations suggest that we look into the geometry of four-dimensional space. Since time does not permit us to do this in full detail, we shall merely open the door into the fourth dimension and leave it slightly ajar. Then in Chapter 6 we shall acquaint you with the simplest four-dimensional figure, namely the four-dimensional cube.

By now you have probably asked yourself some of these questions: "How much sense does it make to talk about this imaginary four-dimensional space?" "To what extent can the analogy with ordinary geometry be exploited to construct a geometry in four-dimensional space?" "How does four-dimensional geometry agree with three-dimensional geometry, and how does it differ?" After careful study, mathematicians have answered these questions as follows: "A four-dimensional geometry can be constructed, and it resembles ordinary three-dimensional geometry in many ways. Moreover, four-dimensional geometry generalizes three-dimensional geometry in much the same way that solid geometry generalizes plane geometry. However, there are very important differences between four-dimensional geometry and ordinary geometry, and these differences are very much like the differences between solid geometry and plane geometry."

15. Motion in the Fourth Dimension

Imagine a two-dimensional world, in which all objects are compelled to lie in some plane and are unable to leave this plane. Suppose you are an inhabitant of this "flatland," and

---

1. Similarly it turns out that the volume of an n-dimensional ball of radius R is $\frac{8}{15} \pi^2 R^5$ if $n = 5$, $\frac{1}{6} \pi^3 R^6$ if $n = 6$ and $\frac{16}{105} \pi^3 R^7$ if $n = 7$.

do not suspect the existence of three-dimensional space which you cannot even imagine. Let flatland be the xy-plane and think of yourself as a point (do not be offended by this simplifying assumption!). Then any closed curve which does not intersect itself and which cannot be penetrated by a point constitutes a "fence" in flatland, as illustrated in Figure 30a for the case of a circular fence. On the other hand, suppose flatland is embedded in three-dimensional space and you have somehow managed to guess the existence of the third dimension. Then you can easily cross the fence by the simple expedient of jumping over it, as shown schematically in Figure 30b.
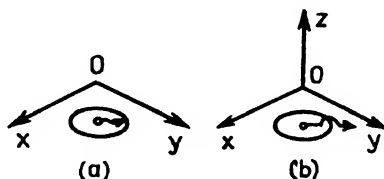

Fig.30. a) The point cannot leave the disk if it stays in flatland; b) The point can leave the disk by moving into the third dimension.

In the same way, suppose you are an inhabitant of the ordinary three-dimensional world (with your permission, we shall again think of you as a point), and suppose you find yourself inside an impenetrable sphere. Then you are clearly doomed to stay inside the sphere forever. However, suppose the sphere is embedded in four-dimensional space, and you have somehow managed to guess the existence of the fourth dimension. Then you can easily escape by simply moving into the fourth dimension and "jumping over the spherical wall."

There is nothing particularly mysterious about all this. A sphere "separates" three-dimensional space (i.e., divides three-dimensional space into two parts), but it does not separate four-dimensional space. This is completely analogous to the fact that a circle separates the plane in which it lies, but does not separate three-dimensional space.

As another example illustrating the same phenomenon, consider the two symmetric figures shown in Figure 31 a, each a square marked with an arrow. Clearly there is no motion in the plane with will make the squares coincide (imagine that the squares can slide over each other). However, if motion in the third dimension is permitted, we can easily perform a rotation carrying either square into the other, as shown in Figure 31 b.
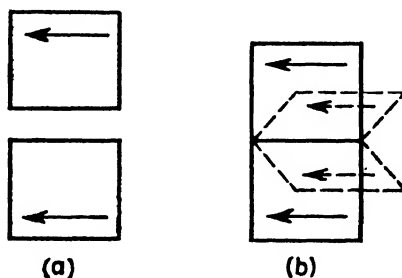

(a)   (b)

Fig. 31. a) The figures cannot be superimposed by making a motion in the plane; b) The figures can be superimposed by making a rotation leaving the plane of the page.

In just the same way, certain symmetric figures in space cannot be brought into coincidence by making any three-dimensional motion. Thus there is no motion carrying a left-handed glove into a right-handed one. However, symmetric figures in space can always be made to coincide if motion in the fourth dimension is permitted, just as symmetric figures in the plane can always be made to coincide if motion in the third dimension is permitted. Thus in "The Plattner Story" by H. G. Wells, the hero Gottfried Plattner takes a trip into four-dimensional space and comes back turned inside out, with his anatomical structure inverted, e.g., with his heart on the right side. This happened because his three-dimensional self somehow got "turned over" during his sojourn in four-dimensional space.

16. Physics and the Fourth Dimension

Four-dimensional geometry turns out to be an extremely useful and well-nigh indispensable tool in contemporary physics. In fact, Einstein's theory of relativity can hardly be stated, much less exploited, without recourse to the apparatus of multidimensional "imaginary" geometry.

Any mathematician would be envious of Minkowski, who after his great success in applying geometry to number theory, went on to use geometric considerations to clarify the difficult mathematics of relativity theory. The key idea of relativity theory is that space and time are inseparably linked concepts. Therefore it is natural to regard the time at which an event occurs as the fourth coordinate of the event, along with the first three coordinates giving the point in space at which the event occurs. The resulting four-dimensional space is called *Minkowski space* or *space-time*, and today every course in relativity theory begins by describing this space. Minkowski's discovery was that the basic formulas of relativity theory, known as the *Lorentz transformation*, become very simple when written in the language of this special four-dimensional space. Thus it was a great boon to modern physics that by the time of the discovery of relativity theory, mathematicians had already developed a convenient, compact and elegant theory of multidimensional geometry, which greatly simplified the solution of a number of problems.

Four-Dimensional Space

17. Coordinate Axes and Planes in Four Dimensions

As promised, we now say something about the geometry of four-dimensional space. In constructing geometry on a line, in a plane or in three-dimensional space, there are two available alternatives: Either we can use intuitive ideas to present the material (this is the customary approach in high school geometry, and hence it would be hard to imagine a geometry textbook without figures), or else, as in the coordinate method, we can do things purely analytically. The first approach is not available for studying four-dimensional space. In fact, four-dimensional space cannot be visualized directly and intuitively, since the space surrounding us has only three dimensions. However, the second alternative is still available. In this approach, a point on a line is defined as a number, a point in the plane is defined as a pair of numbers, and a point in space is defined as a triple of numbers. Therefore it is completely natural to construct a geometry of four-dimensional space by defining a point of this "fictitious space" to be a quadruple of numbers. Then by geometric figures in four-dimensional space we shall mean certain sets of points, just as in the case of ordinary geometry. Let us now make these considerations more precise.

Definition. By a point in four-dimensional space is meant an ordered quadruple of numbers (x, y, z, u).[1]

P

---

1. We say "ordered" since rearranging the numbers in a given quadruple leads to different points (unless the numbers are all the same). Thus the point $(1, -2, 3, 8)$ is different from the point $(3, 1, 8, -2)$.

Now just what are the coordinate axes in four-dimensional space, and how many of them are there? To answer this question we recall the situation in two and three dimensions. In the plane (i.e., in two-dimensional space), a coordinate axis is a set of points such that one coordinate can take any value while the other coordinate equals zero. For example, the x-axis is the set of all points of the form $(x, 0)$, where x is arbitrary. Thus the points $(1, 0)$, $(-3, 0)$ and $(\frac{3}{7}, 0)$ all lie on the x-axis, but not the point $(\frac{1}{2}, 2)$. By the same token, the y-axis is the set of all points of the form $(0, y)$, where y is arbitrary. Similarly, in three-dimensional space there are three coordinate axes:

1. The x-axis, consisting of all points of the form $(x, 0, 0)$ where x is arbitrary;

2. The y-axis, consisting of all points of the form $(0, y, 0)$ where y is arbitrary;

3. The z-axis, consisting of all points of the form $(0, 0, z)$ where z is arbitrary.

But four-dimensional space is the set of all points of the form (x, y, z, u), where x, y, z and u are arbitrary. Therefore it is natural to define a *coordinate axis* in four-dimensional space as a set of points such that one coordinate can take any value while the other three coordinates equal zero. Clearly there are four such coordinate axes:

1. The x-axis, consisting of all points of the form $(x, 0, 0, 0)$ where x is arbitrary;

2. The y-axis, consisting of all points of the form $(0, y, 0, 0)$ where y is arbitrary;

3. The z-axis, consisting of all points of the form $(0, 0, z, 0)$ where z is arbitrary;

4. The u-axis, consisting of all points of the form $(0, 0, 0, u)$ where u is arbitrary.

In three-dimensional space there are *coordinate planes* as well as coordinate axes. These are the planes passing through some pair of coordinate axes. For example, the yz-plane is the plane

passing through the y and z-axes. All told, there are three co-ordinate planes in three-dimensional space:

1. The xy-plane, consisting of all points of the form $(x, y, 0)$ where x and y are arbitrary;

2. The yz-plane, consisting of all points of the form $(0, y, z)$ where y and z are arbitrary;

3. The xz-plane, consisting of all points of the form $(x, 0, z)$ where x and z are arbitrary.

Guided by these considerations, it is natural to define a *co-ordinate plane* in four-dimensional space as a set of points such that two coordinates take arbitrary values while the other two coordinates equal zero. For example, the xz-plane in four-dimensional space is the set of all points of the form $(x, 0, z, 0)$ where x and z are arbitrary. How many such planes are there in all? The answer is six, as we see immediately by writing them all down:

1. The xy-plane, consisting of all points of the form $(x, y, 0, 0)$ where x and y are arbitrary;

2. The xz-plane, consisting of all points of the form $(x, 0, z, 0)$ where x and z are arbitrary;

3. The xu-plane, consisting of all points of the form $(x, 0, 0, u)$ where x and u are arbitrary;

4. The yz-plane, consisting of all points of the form $(0, y, z, 0)$ where y and z are arbitrary;

5. The yu-plane, consisting of all points of the form $(0, y, 0, u)$ where y and u are arbitrary;

6. The zu-plane, consisting of all points of the form $(0, 0, z, u)$ where z and u are arbitrary.

In saying that the value of a coordinate is arbitrary, we allow for the possibility of this value being zero. For example, the point $(5, 0, 0, 0)$ belongs to both the xy-plane and the xu-plane, and to another plane as well (which one?). Moreover, it is easy to see that the yz-plane "passes through" the y-axis in the sense that every point of the y-axis belongs to the yz-plane. In fact, an arbitrary point of the y-axis is of the form $(0, y, 0, 0)$ and

hence belongs to the set of points of the form $(0, y, z, 0)$, i.e., to the *yz*-plane Similarly, the set of points in both the *yz* and the *xz*-planes is just the *z*-axis, consisting of all points of the form $(0, 0, z, 0)$.

To summarize, four-dimensional space contains *six* sets of points analogous to the coordinate planes in three-dimensional space, each consisting of all points such that two coordinates take arbitrary values while the other two coordinates vanish. Each of these coordinate planes "passes through" two coordinate axes. For example, the *yu*-plane passes through the *y* and *u*-axes. On the other hand, there are three coordinate planes passing through each coordinate axis. For example, the *xy*, *xz* and *xu*-planes all pass through the *x*-axis, which is called the *intersection* of the three planes. All six coordinate planes share a single point, namely the origin of coordinates $(0, 0, 0, 0)$.

PROBLEM

What is the intersection of the *xy* and *yz*-planes? Of the *xy* and *zu*-planes?

Thus we see that there is a deep analogy between three-dimensional and four-dimensional space. We can make a schematic drawing giving some intuitive idea of the positions of the coordinate axes and planes in four-dimensional space. This is done in Figure 32, where the coordinate planes are
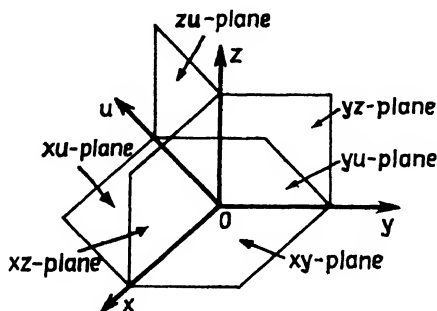

Fig. 32

represented by parallelograms and the coordinate axes by straight lines. The situation is the natural generalization of that shown in Figure 20 for three-dimensional space.

However, in four-dimensional space we can find still other sets of points which might be called coordinate planes. This is really to be expected, since loosely speaking, the line (one-dimensional space) has only an origin, the plane (two-dimensional space) has both an origin and coordinate axes, while three-dimensional space has coordinate planes as well as an origin and coordinate axes. Therefore it is entirely natural that four-dimensional space should contain new "three-dimensional coordinate planes." Each of these planes is a set of points such that three of the four coordinates take arbitrary values while the fourth coordinate equals zero. Thus, for example, the xzu-plane is the set of points of the form $(x, 0, z, u)$ where x, z and u are all arbitrary. Clearly, four-dimensional space contains precisely four such three-dimensional planes:

1. The xyz-plane, consisting of all points of the form $(x, y, z, 0)$ where x, y and z are arbitrary;

2. The xyu-plane, consisting of all points of the form $(x, y, 0, u)$ where x, y and u are arbitrary;

3. The xzu-plane, consisting of all points of the form $(x, 0, z, u)$ where x, z and u are arbitrary;

4. The yzu-plane, consisting of all points of the form $(0, y, z, u)$ where $y, z,$ and u are arbitrary.

Each three-dimensional coordinate plane "passes through" the origin of coordinates and three coordinate axes as well, in the sense that the origin and every point of the axes belong to the plane. For example, the xyu-plane passes through the x, y and u-axes. Similarly, each two-dimensional coordinate plane is the intersection of two three-dimensional coordinate planes. For example, the xy-plane is the intersection of the xyz and xyu-planes, i.e., it consists of all points belonging simultaneously to both the xyz and the xyu-planes.

The situation is shown schematically in Figure 33, which

differs from Figure 32 by containing a sketch of the three-dimensional *xyz*-plane (represented by a parallelepiped). The *xyz*-plane clearly contains the *x*, *y* and *z*-axes and the *xy*, *xz* and *yz*-planes.
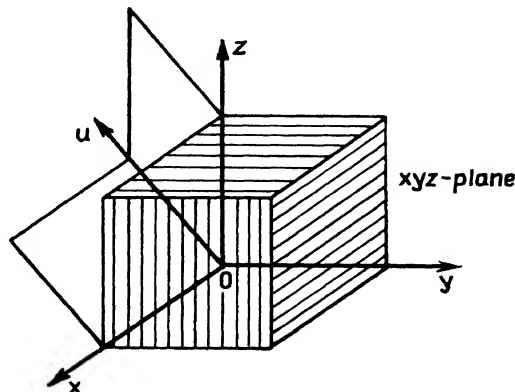

Fig. 33

18. Distance in Four Dimensions

The time has come to think about what is meant by the distance between two points in four-dimensional space. As shown in Secs. 3, 6 and 10, the coordinate method can be used to give a definition of distance that does not involve any geometric models. In fact, the distance between two points $A(x_1)$ and $B(x_2)$ on the line is

$$\varrho\,(A, B) = |x_1 - x_2|$$

or

$$\varrho\,(A, B) = \sqrt{(x_1 - x_2)^2},$$

the distance between two points $A\,(x_1, y_1)$ and $B\,(x_2, y_2)$ in the plane is

$$\varrho\,(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2},$$

while the distance between two points $A\,(x_1, y_1, z_1)$ and $B\,(x_2, y_2, z_2)$ in three-dimensional space is

$$\varrho\,(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

5 Silverman I

The obvious generalization of this formula to the case of four-dimensional space is given by the following

Definition 1. By the distance between two points $A(x_1, y_1, z_1, u_1)$ and $B(x_2, y_2, z_2, u_2)$ in four-dimensional space is meant the quantity

$$\varrho(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 + (u_1 - u_2)^2}.$$

In particular, the distance between the origin $O(0, 0, 0, 0)$ and the point $A(x, y, z, u)$ equals

$$\varrho(O, A) = \sqrt{x^2 + y^2 + z^2 + u^2}.$$

Just as in the case of ordinary three-dimensional geometry, by a figure in four dimensions we mean a set of points in four-dimensional space. Definition 1 suggests the natural way to define a sphere in four dimensions, if as usual we think of a sphere as a set of points all at the same distance from a given point. In fact, choosing the center of the sphere at the origin of coordinates for simplicity, we have

Definition 2. The set of points (x, y, z, u) satisfying the relation

$$x^2 + y^2 + z^2 + u^2 = R^2$$

is called a sphere in four dimensions, more exactly, the sphere of radius R with its center at the origin.

Using Definition 1, we can formulate and solve a number of problems in four-dimensional geometry resembling those encountered in ordinary geometry.

PROBLEMS

1. Prove that the triangle with vertices

$$A(4, 7, -3, 5), \quad B(3, 0, -3, 1), \quad C(-1, 7, -3, 0)$$

is equilateral.

2. Prove that the four points

$$A\,(1, 1, 1, 1),\ B\,(-1, -1, 1, 1),\ C\,(-1, 1, 1, -1),$$
$$D\,(1, -1, 1, -1)$$

are equidistant from each other.

3. Let *A*, *B* and *C* be three distinct points of four-dimensional space and define the angle *ABC* as follows: Using the definition of distance in four dimensions, calculate $\varrho\,(A, B)$, $\varrho\,(B, C)$ and $\varrho\,(A, C)$, i.e., the "side lengths" of the triangle *ABC*. Next construct the triangle *A'B'C'* in the two-dimensional plane, with sides *A'B'*, *B'C'* and *A'C'* of length $\varrho\,(A, B)$, $\varrho\,(B, C)$ and $\varrho\,(A, C)$, respectively. Then the angle *ABC* in four dimensions is defined as the angle *A'B'C'* of the two-dimensional triangle.

Prove that the triangle with vertices

$$A\,(4, 7, -3, 5),\ B\,(3, 0, -3, 1),\ C\,(1, 3, -2, 0)$$

is a right triangle.

4. Find the angles of the triangle figuring in Problem 1.

The Four-Dimensional Cube

19. Definition of the Four-Dimensional Cube
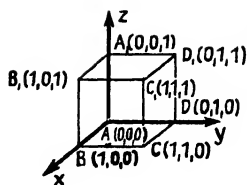
Our last chapter is devoted to a study of the four-dimensional cube. As its name suggests, this figure resembles the familiar three-dimensional cube shown in Figure 34a. There is also a plane figure analogous to the cube, namely the square. The analogy between the two figures becomes transparent if we consider the analytical definitions of the cube and the square. In fact, as you already know from Problem 5, p. 32, a cube can be defined as the set of points (x, y, z) satisfying the inequalities

$$0 \leqslant x \leqslant 1, \ 0 \leqslant y \leqslant 1, \ 0 \leqslant z \leqslant 1. \tag{1}$$

Fig. 34 a

This "arithmetic" definition of the cube does not rely on a figure, but it is completely consistent with the geometric definition of a cube.[1]

The square can also be defined arithmetically, i.e., as the set of points (x, y) satisfying the inequalities

$$0 \leqslant x \leqslant 1, \ 0 \leqslant y \leqslant 1 \tag{2}$$

---

1. Of course, there are other three-dimensional cubes. For example, the set of points defined by the inequalities $-1 \leqslant x \leqslant 1$, $-1 \leqslant y \leqslant 1$, $-1 \leqslant z \leqslant 1$ is also a cube, in fact one very nicely located relative to the coordinate axes, with the origin as its center and the coordinate axes and planes as its axes and planes of symmetry. However, we have chosen the cube defined by (1) as our "basic cube." This cube will sometimes be called the *unit cube*, to distinguish it from other cubes.

(see Fig. 34b). Comparing (1) and (2), we see at once why the square is the "two-dimensional analogue" of the cube. For this reason, the square is sometimes called the "two-dimensional cube."

The analogue of these figures on the line (i.e., in one-dimensional space) is the set of points x satisfying the inequality

$$0 \leqslant x \leqslant 1.$$

Thus the line segment shown in Figure 34c represents a "one-dimensional cube."

We hope that by now you will find the following definition perfectly natural:

Definition. The set of points (x, y, z, u) satisfying the inequalities

$$0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1,\ 0 \leqslant z \leqslant 1,\ 0 \leqslant u \leqslant 1$$

is called a four-dimensional cube.

You mustn't complain that we have not yet drawn a picture of the four-dimensional cube. We will do this later after first analyzing how the cube is "put together" out of its component elements. Do not be surprised to learn that a four-dimensional cube can actually be "drawn"! After all, a three-dimensional cube can be indicated on a plane sheet of paper.
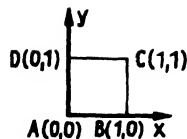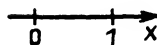

Fig. 34 b

Fig. 34 c

20. Structure of the Four-Dimensional Cube

We begin by considering in turn various "cubes" of different dimensionalities, i.e., the segment, the square and the ordinary cube.

The segment defined by the relations $0 \leqslant x \leqslant 1$ is a very simple figure indeed, about which there is nothing much to say except that its boundary consists of the two points 0 and 1 (the remaining points are called *interior points*). The boundary of the square is more complicated, consisting of four vertices

(points) and four sides (segments). Thus the square has two kinds of boundary elements, namely points and segments. As for the three-dimensional cube, it has three kinds of boundary elements, namely eight vertices (points), twelve edges (segments) and six faces (squares). These facts are summarized in the following table:

Figure	Boundary elements		
	Points (vertices)	Segments (sides, edges)	Squares (faces)
Segment	2	—	—
Square	4	4	—
Cube	8	12	6

This table can be written more compactly by replacing the name of each figure by the number m equal to its dimensionality. Thus $m = 1$ for the segment, $m = 2$ for the square and $m = 3$ for the cube. Similarly, we can replace the name of each boundary element by the number n equal to its dimensionality. Thus $n = 1$ for a side or edge, $n = 2$ for a face, while $n = 0$ for a vertex (a point is regarded as having no dimensions). Our table then becomes

Dimensionality of the cube	Dimensionality of the boundary		
	0	1	2
1	2	—	—
2	4	4	—
3	8	12	6
4			

Our goal is to fill in the fourth row of this table. To do so, we again examine the boundaries of the segment, square and cube, but this time from an analytical (i.e., purely arithmetical) stand-

point, with a view to making the natural generalization to the four-dimensional case.

The boundary of the segment $0 \leqslant x \leqslant 1$ consists of two points (vertices) $x = 0$ and $x = 1$. The boundary of the square $0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1$ contains four vertices

$$x = 0,\ y = 0;\quad x = 0,\ y = 1;\quad x = 1,\ y = 0;\quad x = 1,\ y = 1,$$

i.e., the points $\quad (0, 0),\ (0, 1),\ (1, 0),\ (1, 1).$

The cube $0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1,\ 0 \leqslant z \leqslant 1$ contains eight vertices, obtained by assigning each coordinate of the point (x, y, z) either the value 0 or the value 1. This gives eight points in all:

$$\begin{aligned} &(0, 0, 0),\ (0, 0, 1),\ (0, 1, 0),\ (0, 1, 1), \\ &(1, 0, 0),\ (1, 0, 1),\ (1, 1, 0),\ (1, 1, 1). \end{aligned} \tag{1}$$

Thus, by analogy, we have

Definition 1. By a vertex of the four-dimensional cube $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1,\ 0 \leqslant z \leqslant 1,\ 0 \leqslant u \leqslant 1$ *is meant any point* (x, y, z, u) *such that each coordinate* x, y, z, u *equals either* 0 *or* 1.

There are sixteen such vertices, since there are sixteen distinct quadruples consisting entirely of zeros and ones. In fact, two quadruples can be obtained from each of the triples (1), one by adding zero as the fourth element, the other by adding one as the fourth element. This gives $8 \cdot 2 = 16$ quadruples in all, thereby accounting for all the vertices of the four-dimensional cube.

Next we consider what is meant by an edge of a four-dimensional cube, again relying on analogy. The edges (sides) of a square are determined by the following relations (see Fig. 34 b):

$$0 \leqslant x \leqslant 1,\ y = 0 \text{ (the edge } AB),$$

$$x = 1,\ 0 \leqslant y \leqslant 1 \text{ (the edge } BC),$$

$$0 \leqslant x \leqslant 1,\ y = 1 \text{ (the edge } CD),$$

$$x = 0,\ 0 \leqslant y \leqslant 1 \text{ (the edge } DA).$$

Thus one coordinate of every point belonging to the side of a square has the fixed value 0 or 1, while the other coordinate takes all values between 0 and 1. Similarly, the edges of a three-dimensional cube are given by

$$x = 0, \ y = 0, \ 0 \leqslant z \leqslant 1 \ (\text{the edge } AA_1),$$

$$0 \leqslant x \leqslant 1, \ y = 0, \ z = 1 \ (\text{the edge } A_1B_1),$$

$$x = 1, \ 0 \leqslant y \leqslant 1, \ z = 1 \ (\text{the edge } B_1C_1),$$

and so on (see Fig. 34a).

This suggests

Definition 2. By an edge of the four-dimensional cube $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$, $0 \leqslant z \leqslant 1$, $0 \leqslant u \leqslant 1$ is meant any set of points (x, y, z, u) such that three coordinates x, y, z, u have fixed values equal to 0 or 1 while the other coordinate takes all values between 0 and 1.

For example, the following three sets of points are edges of the four-dimensional cube:

$$x = 0, \ y = 0, \ z = 1, \ 0 \leqslant u \leqslant 1, \tag{2}$$

$$0 \leqslant x \leqslant 1, \ y = 1, \ z = 0, \ u = 1, \tag{3}$$

$$x = 1, \ 0 \leqslant y \leqslant 1, \ z = 0, \ u = 0, \tag{4}$$

etc. To determine the total number of edges of the four-dimensional cube, we have to count the number of sets of relations like (2), (3) and (4). To avoid confusion, we shall do this in a definite order, distinguishing four groups of edges. In the first group, x is chosen as the variable coordinate, ranging between 0 and 1, with y, z and u assigned all possible combinations of the fixed values 0 and 1. But as we already know, there are precisely eight distinct triples consisting of zeros and ones, the same as the number of vertices of the three-dimensional cube [recall (1)]. Hence there are eight edges in the first group, with x as the "free coordinate." By the same token, there are eight edges in the second group, with y as the free coordinate, and

similarly for the third and fourth groups, as shown below, giving the four-dimensional cube a total of $4 \cdot 8 = 32$ edges in all:

First group			Second group			Third group			Fourth group		
$0 \leqslant x \leqslant 1$			$0 \leqslant y \leqslant 1$			$0 \leqslant z \leqslant 1$			$0 \leqslant u \leqslant 1$		
y	z	u	x	z	u	x	y	u	x	y	z
0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	1	0	0	1	0	0	1
0	1	0	0	1	0	0	1	0	0	1	0
0	1	1	0	1	1	0	1	1	0	1	1
1	0	0	1	0	0	1	0	0	1	0	0
1	0	1	1	0	1	1	0	1	1	0	1
1	1	0	1	1	0	1	1	0	1	1	0
1	1	1	1	1	1	1	1	1	1	1	1

Besides vertices and edges, the three-dimensional cube has faces. On each face, one coordinate has a fixed value equal to 0 or 1 while the other coordinates take all values between 0 and 1. For example, the face ABB_1A_1 in Figure 34a is determined by the relations

$$0 \leqslant x \leqslant 1, \ y = 0, \ 0 \leqslant z \leqslant 1.$$

By analogy, in four dimensions we have

Definition 3. By a two-dimensional face[2] of the four-dimensional cube $0 \leqslant x \leqslant 1, \ 0 \leqslant y \leqslant 1, \ 0 \leqslant z \leqslant 1, \ 0 \leqslant u \leqslant 1$ is meant any set of points (x, y, z, u) such that two coordinates x, y, z, u have fixed values equal to 0 or 1 while the other two coordinates take all values between 0 and 1.

For example, the set of points

$$x = 0, \quad 0 \leqslant y \leqslant 1, \quad z = 1, \quad 0 \leqslant u \leqslant 1$$

2. As opposed to a *three-dimensional* face, introduced in Definition 4, p. 63.

is such a two-dimensional face. The four-dimensional cube has $4 \cdot 6 = 24$ two-dimensional faces in all, classified in the following six groups:

First group		Second group		Third group	
$0 \leqslant x \leqslant 1$		$0 \leqslant x \leqslant 1$		$0 \leqslant x \leqslant 1$	
$0 \leqslant y \leqslant 1$		$0 \leqslant z \leqslant 1$		$0 \leqslant u \leqslant 1$	
z	u	y	u	y	z
0	0	0	0	0	0
0	1	0	1	0	1
1	0	1	0	1	0
1	1	1	1	1	1

Fourth group		Fifth group		Sixth group	
$0 \leqslant y \leqslant 1$		$0 \leqslant y \leqslant 1$		$0 \leqslant z \leqslant 1$	
$0 \leqslant z \leqslant 1$		$0 \leqslant u \leqslant 1$		$0 \leqslant u \leqslant 1$	
x	u	x	z	x	y
0	0	0	0	0	0
0	1	0	1	0	1
1	0	1	0	1	0
1	1	1	1	1	1

We can now fill in the fourth row of the table on p. 58, obtaining

Dimensionality of the cube	Dimensionality of the boundary		
	0	1	2
1	2	—	—
2	4	4	—
3	8	12	6
4	16	32	24

However, the table is still not complete, since it should really have another column. In fact, a segment has only one kind of boundary element, a vertex, but a square has edges (sides) as well as vertices, while a three-dimensional cube has (two-dimensional) faces as well as vertices and edges. Thus it must be expected that a four-dimensional cube will have a new kind of boundary element, of dimensionality three, besides vertices, edges and two-dimensional faces.

These considerations lead to

Definition 4. By a three-dimensional face of the four-dimensional cube $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$, $0 \leqslant z \leqslant 1$, $0 \leqslant u \leqslant 1$ *is meant any set of points* (x, y, z, u) *such that one coordinate* x, y, z, u *has a coordinate equal to either* 0 *or* 1 *while the other three coordinates take all values between* 0 *and* 1.

For example, the set of points

$$x = 0, \qquad 0 \leqslant y \leqslant 1 \qquad 0 \leqslant z \leqslant 1 \qquad 0 \leqslant u \leqslant 1$$

is such a three-dimensional face. It is easy to see that the four-dimensional cube has precisely $4 \cdot 2 = 8$ three-dimensional faces, since each of the four coordinates x, y, z, u can take one of the two values 0 or 1 while the other three take all values between 0 and 1.

We can now write the final version of the table on pp. 58 and 62:

Dimensionality of the cube	Dimensionality of the boundary			
	0	1	2	3
1	2	—	—	—
2	4	4	—	—
3	8	12	6	—
4	16	32	24	8

Now take a look at Figure 35, where we have made a drawing of the four-dimensional cube, showing all 16 vertices, 32 edges,
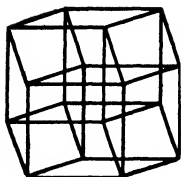
Fig. 35

24 two-dimensional faces (represented by parallelograms) and 8 three-dimensional faces (represented by parallelepipeds). It is clear from the figure which face contains which edge, and so on. The figure is obtained in much the same way as parallel projection[3] is used to represent an ordinary three-dimensional cube on a plane sheet of paper. Thus we first make a model of the three-dimensional figure representing the projection of the four-dimensional cube onto three-dimensional space, and then we sketch this model. If your fingers are nimble enough, you can make such a model out of wooden matches and little balls of modelling clay. (How many matches and balls are needed? How many matches must be stuck into each ball?)

There is another way of getting an intuitive picture of a four-dimensional cube. Suppose you have been asked to send somebody a model of an ordinary three-dimensional cube. Then you can of course use "three-dimensional mail," but this means sending a package, which is doing things the hard way. It is simpler to paste together a cardboard model of the cube, and then unpaste it and send the template as a letter. Such a template of the cube is shown in Figure 36. Since the figure shows the vertices of the cube, it is easy to see how the template is reassembled into a model of the original cube.

In the same way, it is possible to make a "three-dimensional template" of a four-dimensional cube, which turns out to be a figure consisting of eight cubes. Construct such a template or
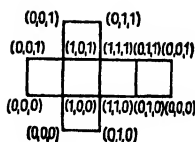
Fig. 36

---

3. Parallel projection of the ordinary cube onto a plane can be accomplished as follows: Make a cube (or rather a cubical frame) out of wire, and observe the shadow cast by the cube on a wall or on a sheet of paper when illuminated by the sun's rays. If the cube is properly placed with respect to the sun's rays, the shadow will resemble the picture of a cube found in geometry books. From a mathematical standpoint, this construction corresponds to drawing a straight line through every point of the cube parallel to a fixed direction (the sun's rays are parallel to each other), but not necessarily perpendicular to the plane. Then the parallel projection of the cube onto the plane is the figure formed by the intersection of the family of straight lines with the plane.

visualize it. Then make a drawing of it, labelling each vertex with the coordinates of the corresponding vertex of the four-dimensional cube.

<center>**PROBLEM**</center>

Write down the relations determining all the three-dimensional faces of the four-dimensional cube.

21. Conclusion

We conclude our primer with another set of problems, testing your knowledge of the four-dimensional cube.

<center>**PROBLEMS**</center>

1. Each edge of the four-dimensional cube is of length 1, just as in the case of the square or the ordinary cube (here by the length of an edge we mean the distance between the vertices lying on the edge).[4] Calculate the distances between the vertices of the four-dimensional cube which do not lie on the same edge.

Hint. Calculate the distance between the vertex $(0, 0, 0, 0)$ and all the other vertices. Since you know the formula for distance in four dimensions (see Sec. 18), the necessary calculations are straightforward.

2. After solving Problem 1, you will find that the vertices of the four-dimensional cube fall into four groups. The vertices of the first group are at distance 1 from $(0, 0, 0, 0)$. Those of the second group are at distance $\sqrt{2}$, those of the third at distance $\sqrt{3}$ and those of the fourth at distance $\sqrt{4} = 2$. How many vertices belong to each group?

3. Of all the vertices of the four-dimensional cube, the distance between $(1, 1, 1, 1)$ and $(0, 0, 0, 0)$ is the largest,

---

4. Thus our cube is called a "unit cube" for good reason.

namely 2. Thus (1, 1, 1, 1) is called the vertex *opposite* (0, 0, 0, 0). and the segment joining (0, 0, 0, 0) and (1, 1, 1, 1) is called the *main diagonal* of the four-dimensional cube. Find the main diagonal and its length for the cubes of different dimensionality.

4. Imagine an ant sitting on the vertex (0, 0, 0) of a three-dimensional cube made out of wire. Then the ant can crawl from one vertex to another along the edges. To arrive at the vertex (1, 1, 1), the ant must crawl along three edges, and hence (1, 1, 1) is called a *third-order vertex*. Similarly, the path from (0, 0, 0) to the vertex (0, 1, 1) consists of two edges, and hence a vertex like (0, 1, 1) is called a *second-order vertex*. The cube also has *first-order vertices*, namely the points (0, 0, 1), (0, 1, 0) and (1, 0, 0), which our ant can reach from (0, 0, 0) by crawling along one edge. Write the coordinates of the three second-order vertices of the cube.

Actually there are two paths consisting of two edges and joining (0, 0, 0) to each second-order vertex. For example, the vertex (0, 1, 1) can be reached from (0, 0, 0) by going through either of the vertices (0, 0, 1) and (0, 1, 0). How many paths consisting of three edges are there joining opposite vertices?

5. Consider the four-dimensional cube

$$-1 \leqslant x \leqslant 1, \quad -1 \leqslant y \leqslant 1, \quad -1 \leqslant z \leqslant 1, \quad -1 \leqslant u \leqslant 1$$

centered at the origin. Find the distance from the vertex (1, 1, 1, 1) to all the other vertices of the cube. Which vertices are first-order vertices relative to the vertex (1, 1, 1, 1), i.e., which vertices can be reached from (1, 1, 1, 1) by going along a single edge? Which vertices are second-order vertices? Third-order? Fourth-order?

6. How many paths consisting of four edges are there going from the vertex (0, 0, 0, 0) of the four-dimensional unit cube to the opposite vertex (1, 1, 1, 1)? Write each path down in detail, showing the vertices it goes through consecutively.

Index